

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 15:44:49 ; Search time 21.9534 Seconds
(without alignments)
2602.858 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSGVATFVLIQEEFDR.....RVAIAGAPVTLIMFYDTCYT 360

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_19Jun03.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1933	100.0	360	22	Human DPP8 244Glu-
2	1836.5	95.0	882	22	Human DPP8 244Glu-
3	1836.5	95.0	882	23	Human dipeptidyl p
4	1836.5	95.0	882	23	Human dipeptidyl p
5	1836.5	95.0	882	23	Human DPPIV relate
6	1836.5	95.0	882	23	Human protease PRT
7	1836.5	95.0	882	24	Amino acid sequenc
8	1528	79.0	632	22	Human serine prote
9	1254.5	64.9	724	23	Novel human protei

10	1254.5	64.9	782	23	AB897361
11	1220.5	63.1	830	23	AAE24171
12	1220.5	63.1	863	23	ABG61592
13	1220.5	63.1	892	23	AB898134
14	1220.5	63.1	892	23	ABG61602
15	1220.5	63.1	892	23	ABG61604
16	1220.5	63.1	969	23	AAE24168
17	1204.5	62.3	847	23	AAE23875
18	1204.5	62.3	869	23	AAE24169
19	1201	62.1	879	23	ABG61607
20	1201	62.1	879	23	ABG61608
21	1199.5	62.1	658	23	ABG61600
22	1199.5	62.1	661	23	ABG61596
23	1199.5	62.1	690	23	ABG61594
24	1190	61.6	580	23	AAE14337
25	1183.5	61.2	683	22	AA440510
26	1183.5	61.2	737	22	AA440510
27	1132.5	58.6	832	23	ABG61605
28	1132.5	58.6	832	23	ABG61606
29	1113	57.6	819	23	ABG61609
30	1113	57.6	819	23	ABG61610
31	1086.5	56.2	310	22	AA847188
32	1086.5	56.2	310	23	AB808994
33	984	50.9	497	23	ABG64845
34	984	50.9	497	23	AA961659
35	980	50.7	497	23	ABG64844
36	980	50.7	497	23	AA961692
37	944	48.8	508	21	AA842928
38	931.5	48.2	613	23	ABG61601
39	860.5	44.5	465	22	AA847189
40	967	34.7	250	23	AB899949
41	722	37.4	1042	22	AB860137
42	722	37.4	1102	22	AB862029
43	654.5	33.9	720	21	AA841626
44	505	26.1	358	23	ABG61597
45	403.5	20.9	732	21	AA818507

ALIGNMENTS

RESULT 1
ID AAB47190 standard; Protein, 360 AA.
XX AAB47190;
AC
XX 29-JUN-2001 (first entry)
DT
XX Human DPP8 244Glu-341Ile+515Val-776Thr.
DE
XX Human, dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
KW growth hormone deficiency; glucose level; mucosal regeneration;
KW non-insulin dependent diabetes mellitus; glucose intolerance;
KW immunosuppression.
XX Homo sapiens.
OS
XX WO200119866-A1.
PN
XX 22-MAR-2001.
PD
XX 11-SEP-2000; 2000WO-AU01085.
PF
XX 10-SEP-1999; 99AU-0002762.
PR 18-FEB-2000; 2000AU-0005709.
XX (UNSY P UNITV SYDNEY.
PA Abbot CA, Gorell MD;
XX WPI, 2001-281520/29.
DR

Novel human protei
Human dipeptidyl p
Human DPPIV relate
Human PMW Incyte
Human DPP-2 splic
Human DPP-2 splic
Human dipeptidyl p
Murine dipeptidyl
Alternative versio
Human DPP-2 splic
Human DPP-2 splic
Human DPP-1 splic
Human DPP-1 splic
Human DPP-1 splic
Human protease PRT
Human polypeptide
Human DPP-2 splic
Human DPP-2 splic
Human DPP-2 splic
Human DPP8 524phe-
Human dipeptidyl p
Human albumin fusi
Human secreted pro
Human albumin fusi
Human secreted pro
Human ORFX ORF692
Human DPP8 318Thr-
Dipeptidyl peptida
Drosophila melanog
Human ORFX ORF1390
Human DPP-1 splic
Amino acid sequenc

Query Match	Best Local Similarity	Score	DB	Length
Matches 360; Conservative	100.0%;	1933;	DB 22;	360;
	100.0%;	Pred. No. 5.3e-198;		
	0;	Mismatches 0;	Indels 0;	Gaps 0;
Sequence 360 AA;				
DR	N-PSDB; AAC05697.			
XX				
PT	New human dipeptidyl aminopeptidase (DPP8) useful for cleaving			
PT	substrates, identifying inhibitors of DPP8 catalytic activity which			
PT	have therapeutic uses, and for detecting activated T cells			
PS	Claim 5; Page 74-75; 78pp; English.			
XX				
CC	The sequences given in AAB47187-90 represent fragments of human			
CC	dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for			
CC	H-Ala-Pro-PNA, H-Gly-Pro-PNA and H-Arg-Pro-PNA. Therefore, it is a			
CC	prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable			
CC	of hydrolysing the peptide bond C-terminal to Pro in each of these			
CC	compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for			
CC	cleaving a substrate, and for detecting an activated T cell which			
CC	involves measuring the level of DPP8 gene expression in a T cell. The			
CC	level of DPP8 expression is detected by detecting the amount of DPP8			
CC	RNA in the cell. It is also useful for identifying a molecule capable			
CC	of inhibiting the cleavage of the substrate by DPP8. Molecules			
CC	identified as inhibiting DPP8 catalytic activity may be useful for			
CC	treating diarrhoea, growth hormone deficiency, lowering glucose levels			
CC	in non-insulin dependent diabetes mellitus and other disorders			
CC	involving glucose intolerance, enhancing mucosal regeneration and			
CC	as immunosuppressants.			
XX				
XX				
YY	1 EEDARSAAVAFVLOEEDFRDSRGYMCCKAETTPSGKILRLIVENESEVEIITHTSP 60			
DB	1 EEDARSAAVAFVLOEEDFRDSRGYMCCKAETTPSGKILRLIVENESEVEIITHTSP 60			
QY	61 MLETRRADSFRRPKTGTPANPKVTFKMSIMIDAEGRITIVDEVRLVYEGTKDSPLEHNL 120			
DB	61 MLETRRADSFRRPKTGTPANPKVTFKMSIMIDAEGRITIVDEVRLVYEGTKDSPLEHNL 120			
QY	121 YVSVYVNGGEVRLTRDGRYSHSCCISQCHDPFISYSNQKNPHCVSLTKLSSPEDDPCK 180			
DB	121 YVSVYVNGGEVRLTRDGRYSHSCCISQCHDPFISYSNQKNPHCVSLTKLSSPEDDPCK 180			
QY	181 TKEFPAATILDSAGPLPDYTPPEIFSFESSTGFTLLGYMKYKHDLPQKKYPTVLFIYGP 240			
DB	181 TKEFPAATILDSAGPLPDYTPPEIFSFESSTGFTLLGYMKYKHDLPQKKYPTVLFIYGP 240			
QY	241 QVQLVNNRFRKGYKRYRLNTLASLGVVVVYVINDRSGCHGKLFEGAFKRYMGQIEIDDOVE 300			
DB	241 QVQLVNNRFRKGYKRYRLNTLASLGVVVVYVINDRSGCHGKLFEGAFKRYMGQIEIDDOVE 300			
QY	301 GLQYLASRYDFIDLRVGIHGWSYGYSLMALMQRSDIFRVAIGAPVTLMIFFDTGYT 360			
DB	301 GLQYLASRYDFIDLRVGIHGWSYGYSLMALMQRSDIFRVAIGAPVTLMIFFDTGYT 360			
RESULT 2				
AAB47187				
ID	AAB47187 standard; Protein; 882 AA.			
AC	AAB47187;			
XX				
XX	29-JUN-2001 (first entry)			
DT				
XX				
XX	Human DPP8.			
DE				
XX				
XX	Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;			
KM	dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;			
KM	growth hormone deficiency; glucose level; mucosal regeneration;			
KW	non-insulin dependent diabetes mellitus; glucose intolerance;			
KW	immunosuppression.			
XX				
OS	Homo sapiens			

[illegible]

QY 128 PGEVTRLDRGYSHCCTISOHCDPFIISKYNOKNPHCVSLYKLSPEDDPTCKTKEFMAT 187
 DB 544 PGEVTRLDRGYSHCCTISOHCDPFIISKYNOKNPHCVSLYKLSPEDDPTCKTKEFMAT 603
 QY 188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPGKKYPTVLFTYGGPOVOLVNN 247
 DB 604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPGKKYPTVLFTYGGPOVOLVNN 663
 QY 248 RFGKVKYFRLNTLASLGYVVVVVIDNRGSGHGLKEFGAFKXKMGQIEIDDQVEGLQYLAS 307
 DB 664 RFGKVKYFRLNTLASLGYVVVVVIDNRGSGHGLKEFGAFKXKMGQIEIDDQVEGLQYLAS 723
 QY 308 RYDFIDLDVRGIGHGMSYGGYLSLMAIMQORSIDIFRYAIGAPVTLMIFYDTGYT 360
 DB 724 RYDFIDLDVRGIGHGMSYGGYLSLMAIMQORSIDIFRYAIGAPVTLMIFYDTGYT 776

RESULT 3

ID AAE24170 standard; Protein; 882 AA.

AAE24170;

23-SEP-2002 (first entry)

Human dipeptidyl peptidase 8 (DPP8) protein.

Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
 autoimmunity; human immunodeficiency virus; HIV infection; cytostatic;
 graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
 antiviral; enzyme.

OS Homo sapiens.

PN WC0200234900-A1.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-AU01388.

XX 27-OCT-2000; 2000AU-0001078.

XX (UNSY) UNIV SYDNEY.

XX Abbot CA, Gorrell MD;

XX WPI; 2002-454646/48.

XX N-PSDB; AAD38956.

PT New dipeptidyl peptidase (DPP) peptidases, useful for screening
 PT inhibitors of DPP catalytic activity, which may be employed to treat
 PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 PT rejection and HIV infection -

PS Example; Fig 1; 91pp; English.

CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
 CC polymucloides encoding such proteins. The DPP peptidases are useful for
 CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
 CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 CC rejection and HIV (human immunodeficiency virus) infection. The present
 CC sequence is human DPP8 protein.

XX Sequence 882 AA;

Query Match 95.0%; Score 1836.5; DB 23; Length 882;

Best Local Similarity 67.5%; Pred. No. 4.7e-187; Indels 173; Gaps 1;

QY 1 EEDARSAGVATFVLQEEFDRYSGYWCPCAKETTPSGKILRIIVLEENDESEVEIIHTSP 60
 DB 244 EEDARSAGVATFVLQEEFDRYSGYWCPCAKETTPSGKILRIIVLEENDESEVEIIHTSP 303

QY 61 MLETRADSFYRKYTGTAHPKVTFMSEIMIDAEGRIL----- 98
 DB 304 MLETRADSFYRKYTGTAHPKVTFMSEIMIDAEGRILIDVDKELIQPEILFEGVEYIA 363
 QY 99 ----- 98
 DB 364 RAGWTEGKRAMSILDRSQTRLQVLISPELIFVEDDVMERQRLIESVPSVTPLLIY 423
 QY 99 ----- 98
 DB 424 EETDIDMINIHDFHVPQSHHEEIEFIIPASECKTGFRHLKYITSLIKSKYRRSGGLP 483
 QY 99 -----VDEVRLVYFEGTDSPLEHLLYVSYNN 127
 DB 484 APSDEKCPIDKEIATSGEWEVLGRHGSNIQDEVRRLVYFEGTDSPLEHLLYVSYNN 543
 QY 128 PGEVTRLDRGYSHCCTISOHCDPFIISKYNOKNPHCVSLYKLSPEDDPTCKTKEFMAT 187
 DB 544 PGEVTRLDRGYSHCCTISOHCDPFIISKYNOKNPHCVSLYKLSPEDDPTCKTKEFMAT 603
 QY 188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPGKKYPTVLFTYGGPOVOLVNN 247
 DB 604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPGKKYPTVLFTYGGPOVOLVNN 663
 QY 248 RFGKVKYFRLNTLASLGYVVVVVIDNRGSGHGLKEFGAFKXKMGQIEIDDQVEGLQYLAS 307
 DB 664 RFGKVKYFRLNTLASLGYVVVVVIDNRGSGHGLKEFGAFKXKMGQIEIDDQVEGLQYLAS 723
 QY 308 RYDFIDLDVRGIGHGMSYGGYLSLMAIMQORSIDIFRYAIGAPVTLMIFYDTGYT 360
 DB 724 RYDFIDLDVRGIGHGMSYGGYLSLMAIMQORSIDIFRYAIGAPVTLMIFYDTGYT 776

RESULT 4

ID ABG61591 standard; Protein; 882 AA.

ABG61591;

12-AUG-2002 (first entry)

Human DPPIV related serine protease DPP-1.

Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
 DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 heart failure; hypertension; urinary retention; osteoporosis; cancer;
 ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 dyskinesia; reproductive disorder; inflammatory disorder;
 metabolic disorder.

OS Homo sapiens.

PN WC0200231134-A2.

XX 18-APR-2002.

XX 12-OCT-2001; 2001WO-US1874.

XX 12-OCT-2000; 2000US-240117P.

XX (FERR) FERRING BV.

XX Qi S, Aktinsanya KO, Riviere PJ, Junien J;

XX WPI; 2002-444178/47.

XX N-PSDB; ABK83322.

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain

XX Claim 17; Fig 1; 113pp; English.

XX The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPP-IV proteins.

XX Sequence 882 AA:

Query Match 95.0%; Score 1836.5; DB 23; Length 882;
 Best Local Similarity 67.5%; Pred. No. 4.7e-187;
 Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

QY 1 EEDASAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEIIVHTSP 60
 DB 244 EEDASAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEIIVHTSP 303

QY 61 MLETRRADSPRYPKGTANPKVTFKMSIIMDAEGRII----- 98
 DB 304 MLETRRADSPRYPKGTANPKVTFKMSIIMDAEGRIIIVIDKELIOPPELLFEGVEYIA 363

QY 99 ----- 98
 DB 364 RAGWTEGKYAMSLIDRSQTRLQVLISPELLFIVEDDWMERQRLIESVPSVPLIITY 423

QY 99 ----- 98
 DB 424 EETDIWINIHDIHFVFPQSHHEIEIFASECKTGFRHLKYKITSILKSKYKRSGGULP 483

QY 99 -----VDEVRILYFPGTKDSPLEHLLVYSYVN 127
 DB 484 APSDECPKIEEIIATSGEMEVILGRHGSNIQVDEVRLVYFEGTKDSPLEHLLVYSYVN 543

QY 128 PGEVRLTDRGSHSCCTISOHCDPFISSKSNOKNHCVSILYKSSPEDPTCKTEPMAT 187
 DB 544 PGEVRLTDRGSHSCCTISOHCDPFISSKSNOKNHCVSILYKSSPEDPTCKTEPMAT 603

QY 188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFIYGGPOVQLVNN 247
 DB 604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFIYGGPOVQLVNN 663

QY 248 RFKGYKYYRRLNTLASIGYVVVINDNGSCHRLKTEGAKRYMGQIEIDDOVEGIQYLAS 307
 DB 664 RFKGYKYYRRLNTLASIGYVVVINDNGSCHRLKTEGAKRYMGQIEIDDOVEGIQYLAS 723

QY 308 RVDPLIDLVGIGHMSYGGYLSLMLMORSDIFRVAIAGAPVTLMFYDTGYT 360
 DB 724 RVDPLIDLVGIGHMSYGGYLSLMLMORSDIFRVAIAGAPVTLMFYDTGYT 776

RESULT 5
 AAU74749
 ID AAU74749 standard; Protein; 882 AA.
 XX AAU74749;
 AC
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Human protease PRTS-9 protein sequence.
 XX
 KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
 KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;

KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
 KW cell proliferative disorder; developmental disorder; epilepsy;
 KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
 KW reproductive disorder; endometriosis.
 OS Homo sapiens.
 PN WO200196468-A2.
 PD 27-DEC-2001.
 PF 13-JUN-2001; 2001WO-US19178.
 PR 16-JUN-2000; 2000US-212336P.
 PR 22-JUN-2000; 2000US-213955P.
 PR 29-JUN-2000; 2000US-215396P.
 PR 07-JUL-2000; 2000US-216821P.
 PR 14-JUL-2000; 2000US-218946P.
 XX
 PA (INCYTE GENOMICS INC.
 XX Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
 PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
 PI Walla NK, Yao MG, Lu DM, Patterson C, Tang YR, Walsh RT;
 PI Azimzal Y, Lu Y, Rankumar J, Xu Y, Reddy R, Das D, Kearney L;
 PI Kallik Da,
 DR WPI; 2002-090437/12.
 DR N-PSDB; ABK12892.
 XX
 PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
 PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
 PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
 PT proliferative (e.g. cancer) disorders -
 PS
 XX
 PS Claim 1; Page 140-142; 177pp; English.

The present invention relates to twenty one new human proteases,
 referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
 polypeptides of the invention are useful in the diagnosis, treatment and
 prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
 Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
 myocardial infarction, autoimmune/inflammatory e.g. acquired
 immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
 CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
 CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
 CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
 CC endometriosis disorders. Numerous other examples of each disorder are
 CC given in the specification. The present protein sequence represents
 the human protease PRTS-9 protein of the invention.

Sequence 882 AA:

Query Match 95.0%; Score 1836.5; DB 23; Length 882;
 Best Local Similarity 67.5%; Pred. No. 4.7e-187;
 Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

QY 1 EEDASAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEIIVHTSP 60
 DB 244 EEDASAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEIIVHTSP 303

QY 61 MLETRRADSPRYPKGTANPKVTFKMSIIMDAEGRII----- 98
 DB 304 MLETRRADSPRYPKGTANPKVTFKMSIIMDAEGRIIIVIDKELIOPPELLFEGVEYIA 363

QY 99 ----- 98
 DB 364 RAGWTEGKYAMSLIDRSQTRLQVLISPELLFIVEDDWMERQRLIESVPSVPLIITY 423

QY 99 ----- 98
 DB 424 EETDIWINIHDIHFVFPQSHHEIEIFASECKTGFRHLKYKITSILKSKYKRSGGULP 483

QY 99 -----VDEVRLVYFEGTKDSPLEHLLVYVSYN 127
 Db 484 APSDKCPKIEKIATITSGWEVLGRHGSNIQVDEVARLVYFEGTKDSPLEHLLVYVSYN 543
 QY 128 PGEVTRLDRGYSHSCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMAT 187
 Db 544 PGEVTRLDRGYSHSCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMAT 603
 QY 188 IIDSAGPLPDYPPPELFFSFESTTGFTLYGMLYKPHDLPQKKYPTVLYFGGPQVQLVNN 247
 Db 604 IIDSAGPLPDYPPPELFFSFESTTGFTLYGMLYKPHDLPQKKYPTVLYFGGPQVQLVNN 663
 QY 248 RFKGVKYPRLNTLASIGYVVVVINDRGSCHRGLKEFGAKFYKMGQIEIDDOVEGLQYLAS 307
 Db 664 RFKGVKYPRLNTLASIGYVVVVINDRGSCHRGLKEFGAKFYKMGQIEIDDOVEGLQYLAS 723
 QY 308 RYDFIDLDRVGIHGWSYGYLSLMLMQRSDIFRVAIAGAPVTLMWTFYDTGYT 360
 Db 724 RYDFIDLDRVGIHGWSYGYLSLMLMQRSDIFRVAIAGAPVTLMWTFYDTGYT 776
 RESULT 6
 AAG78415 ID AAG78415 standard; Protein; 882 AA.
 AC AAG78415;
 DT 12-APR-2002 (first entry)
 DE Amino acid sequence of 21953 human prolyl oligopeptidase.
 XX
 XX
 XX 21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
 KM cancer; cardiovascular disease; autoimmune disease; atopic allergy;
 KM neuronal disorder; vascular disorder; prostate disorder; cytostatic;
 KM antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
 KM diabetes mellitus; arthritis; multiple sclerosis; asthma;
 KM Grave's disease; neuronal disorder; demyelinating disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200179473-A2.
 XX
 PD 25-OCT-2001.
 XX
 PP 11-APR-2001; 2001WO-US40483.
 XX
 PR 18-APR-2000; 2000US-197508P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Meyers RA, Williamson M;
 XX
 XX WPI; 2002-034353/04.
 DR N-PSDB; AAH99934.
 PT
 PT New polypeptides 21953, member of human prolyl oligopeptidase family,
 PT useful as diagnostic targets and therapeutic agents for controlling
 PT cancer, lymphoma and leukemia
 PS Claim 1; Page 102-103; 121pp; English.
 XX
 XX This invention relates to an isolated 21953 human prolyl
 CC oligopeptidase. Which is cytosolic, antidiabetic, antiarthritic,
 CC neuroprotective, antihypertensive, dermatological, antipsoriatic,
 CC antiasthmatic, ophthalmological, antiinflammatory, nootropic,
 CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
 CC antidiabetic, cardiant, antiatherosclerotic, anorectic and
 CC metabolic in its action. Uses include gene therapy, expression or
 CC activity of 21953 protein modulator, it is useful for identifying a
 CC compound which binds to it and can be used in preventing, treating
 CC or detecting a cellular proliferative or differentiative disorder.
 CC The 21953 molecules can act as novel diagnostic targets and therapeutic
 CC agents for controlling disorders associated with the aberrant activity

CC or degradation of peptide hormones e.g., disorders associated with cell
 CC differentiation and proliferation such as cancer, immune function,
 CC reproductive, neurological and cardiovascular function. The 21953
 CC molecules are thus useful for treating and preventing cellular
 CC proliferative and differentiative disorders, haematopoietic neoplastic
 CC disorders, immune disorders such as autoimmune diseases, diabetes
 CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
 CC neuronal disorders, demyelinating diseases, vascular disorders and
 CC metabolism or pain disorders. This sequence represents the amino
 CC acid sequence of 21953 human prolyl oligopeptidase.
 SQ Sequence 882 AA;
 Query Match 95.0%; Score 1836.5; DB 23; Length 882;
 Best Local Similarity 67.5%; Pred. No. 4.7e-187;
 Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
 QY 1 EEDARSAGVATVLOEFPDRYSGYWWCPAETTPSGKTLRLYEENDESEVLIHVTSP 60
 Db 244 EEDARSAGVATVLOEFPDRYSGYWWCPAETTPSGKTLRLYEENDESEVLIHVTSP 303
 QY 61 MLETRRADSFYRYPKTGTANPKVTFKMSIIMDAEGR11----- 98
 Db 304 MLETRRADSFYRYPKTGTANPKVTFKMSIIMDAEGR11IDVDEKELQPEELIFEGVEYIA 363
 QY 99 ----- 98
 Db 364 RAGWTPGCKYAMSILLDRSQTRLQIVLISPDLFIPVEDDMERQRLIESVPSVTPLIY 423
 QY 99 ----- 98
 Db 424 ERTTIDWINIHDI FHVFPQSHHEBIEFIPASCKTGFRHLKYITSILKSKYKRSGLP 483
 QY 99 -----VDEVRLVYFEGTKDSPLEHLLVYVSYN 127
 Db 484 APSDKCPKIEKIATITSGWEVLGRHGSNIQVDEVARLVYFEGTKDSPLEHLLVYVSYN 543
 QY 128 PGEVTRLDRGYSHSCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMAT 187
 Db 544 PGEVTRLDRGYSHSCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMAT 603
 QY 188 IIDSAGPLPDYPPPELFFSFESTTGFTLYGMLYKPHDLPQKKYPTVLYFGGPQVQLVNN 247
 Db 604 IIDSAGPLPDYPPPELFFSFESTTGFTLYGMLYKPHDLPQKKYPTVLYFGGPQVQLVNN 663
 QY 248 RFKGVKYPRLNTLASIGYVVVVINDRGSCHRGLKEFGAKFYKMGQIEIDDOVEGLQYLAS 307
 Db 664 RFKGVKYPRLNTLASIGYVVVVINDRGSCHRGLKEFGAKFYKMGQIEIDDOVEGLQYLAS 723
 QY 308 RYDFIDLDRVGIHGWSYGYLSLMLMQRSDIFRVAIAGAPVTLMWTFYDTGYT 360
 Db 724 RYDFIDLDRVGIHGWSYGYLSLMLMQRSDIFRVAIAGAPVTLMWTFYDTGYT 776
 RESULT 7
 ABU07720 ID ABU07720 standard; Protein; 882 AA.
 XX
 XX ABU07720;
 AC
 DT 19-MAY-2003 (first entry)
 DE Human serine protease HIPH0M46.
 XX
 XX Human; enzyme; HIPH0M46; serine protease; gene therapy; osteoarthritis;
 KM serine protease activity modulation; dipeptidyl peptidase activity;
 KM musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
 KM Alzheimer's disease; paraspinal clear palsy; Huntington's disease;
 KM amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
 KM irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
 KM haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
 KM colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
 KM multiple sclerosis.

```

XX OS Homo sapiens.
XX Key Location/Qualifiers
XX Region 259..260
XX /note= "Paired glutamates of the beta propeller domain"
FT Active-site 739
FT /label= Catalytic_serine_residue
FT Active-site 817
FT /label= Catalytic_aspartate_residue
FT Active-site 849
FT /label= Catalytic_histidine_residue
XX GB2374869-A.
XX 30-OCT-2002.
XX 22-JAN-2002; 2002GB-0001404.
XX 23-JAN-2001; 2001GB-0001760.
XX (GLAXO ) GLAXO GROUP LTD.
XX Edbrooke MR, Lewis AP;
XX MPI; 2003-150703/15.
XX N-PSDB; ABX12255.
XX
XX Identifying modulators of serine protease activity useful for treating
XX musculoskeletal diseases, by contacting cell expressing a novel serine
XX protease polypeptide with a compound and monitoring serine protease
XX activity
XX
XX Claim 10; Page 26-29; 38pp; English.
XX
XX The invention relates to a method of identifying a substance that
XX modulates serine protease activity, comprising contacting a cell such as
XX a neuronal cell, lung cell, intestinal cell or a cell infected with a
XX virus, expressing a serine protease polypeptide (HHPHM 46), or its
XX variant having dipeptidyl peptidase activity, or a serine protease
XX isolated from the cell with a test substance and monitoring for serine
XX protease activity. The method is useful for identifying a substance that
XX modulates serine protease activity. A modulator of the serine protease is
XX useful in the manufacture of a medicament for treatment or prophylaxis of
XX a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
XX infection, Alzheimer's disease, paraneoplastic myotonic
XX dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
XX Additional disease that may be treated using modulators of the serine
XX protease include malabsorption syndromes, irritable bowel syndrome, lung
XX disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
XX rectal polyps, small bowel tumours, colorectal tumours, anaemia,
XX dystexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
XX sclerosis. The present sequence represents the amino acid sequence of the
XX human serine protease HHPHM46.
XX
XX Sequence 882 AA:
XX
XX Query Match 95.0%; Score 1836.5; DB 24; Length 882;
XX Best Local Similarity 67.5%; Pred. No. 4.7e-187; Indels 173; Gaps 1;
XX Matches 360; Conservative 0; Mismatches 0;

```

```

QY 99 ----- 98
Db 424 EETDIDWINTHIDIRHVPQSHBEIEPIFASECKTGFRHLKYITSLKESKXKSSGLP 483
QY 99 ----- VDEVRLVYFEGTKDSPLEHLLYVVSYN 127
Db 484 APSDFKCPKEIAITSGEMEVLRHGSNIQVDEVRRLVYFEGTKDSPLEHLLYVVSYN 543
QY 128 PGEVTRLTDRGYSHSCISQCHDFFISKYSNQKPHCVSLYKLSPEDDPTKTEFMAT 187
Db 544 PGEVTRLTDRGYSHSCISQCHDFFISKYSNQKPHCVSLYKLSPEDDPTKTEFMAT 603
QY 188 ILDSAGLPDVTPTPEIFSESTTGFTLGMVYKPHDLQPGKKYPTVLFIYGSPQVLVN 247
Db 604 ILDSAGLPDVTPTPEIFSESTTGFTLGMVYKPHDLQPGKKYPTVLFIYGSPQVLVN 663
QY 248 RFGVKYFRLNTLASLGYVWVVDNRGSGHGLKEGAFKXMGQIEIDDQVEGLQYLAS 307
Db 664 RFGVKYFRLNTLASLGYVWVVDNRGSGHGLKEGAFKXMGQIEIDDQVEGLQYLAS 723
QY 308 RYDFIDLRVGIHGWSYGGYISLWALMQRSDIFRVAIGAPVTLMIFFDTGTT 360
Db 724 RYDFIDLRVGIHGWSYGGYISLWALMQRSDIFRVAIGAPVTLMIFFDTGTT 776

RESULT 8
AAB93565
ID AAB93565 standard; Protein; 632 AA.
AC AAB93565;
XX 26-JUN-2001 (first entry)
DT Human protein sequence SEQ ID NO:12964.
XX Human
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EPI074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX MPI; 2001-318749/34.
XX
XX Claim 8; SEQ ID 12964; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification, where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end

```

CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX SQ Sequence 632 AA;

Query Match 79.0%; Score 1528; DB 22; Length 632;
Best Local Similarity 57.8%; Pred. No. 3.2e-154; Indels 224; Gaps 2;
Matches 308; Conservative 0; Mismatches 1;

QY 1 EEDARSAGVATFVLQEFPRYSGYWCPKAEETPSGKILRIIYENDESEVEIIHVTSP 60
DB 45 EEDARSAGVATFVLQEFPRYSGYWCPKAEETPSGKILRIIYENDESEVEIIHVTSP 104
QY 61 METRRADSFRRYKGTANPKVTFKMSIIMDAEGRII----- 98
DB 105 METRRADSFRRYKGTANPKVTFKMSIIMDAEGRIIIVDKELIQPEILFEGVEYIA 164
QY 99 ----- 98
DB 165 RAGWTPBCKYAMSILDRSQTRLQIVLISPFLIPVEDDYMERQRLIESVPSVTPLIY 224
QY 99 ----- 98
DB 225 EETTDIWINIHDFHVPQSHBEIEFIFASECKTGRRLHYKTTSLIKESKYRSSGGLP 244
QY 99 -----VDEVRRLVYFEGTKDSPLEHLLVYVSYVN 127
DB 285 APSDFKCPIKEEIAITSGEMEVILGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLVYVSYVN 344
QY 128 PGCVTRLTDRGVSHSCCISQHCDFPISKYSNQNPKCVSLYKLSPEDDPTCKTKERWAT 187
DB 345 PGCVTRLTDRGVSHSCCISQHCDFPISKYSNQNPKCVSLYKLSPEDDPTCKTKERWAT 404
QY 188 ILDSAGPLDPYTPPELIFSFESTTGFTLYGMLYKPHLOPKKYPVLFYIGSPQVQLVNN 247
DB 405 ILDSAGPLDPYTPPELIFSFESTTGFTLYGMLYKPHLOPKKYPVLFYIGSPQVQLVNN 464
QY 248 RFGKGVYFRNLNTLASLIGYVVVVDNRGSGHRLKFGAFKYYKMGQIEIDDQVEGLQYLAS 307
DB 465 RFGKGVYFRNLNTLASLIGYVVVVDNRGSGHRLKFGAFKYYKMGQIEIDDQVEGLQYLAS 507
QY 308 RYDFIDLDVGIHGMVSYGYLSLMMQSRDIFRVAIAGAPVTLMTFYDTGYT 360
DB 508 -----VALIAGAPVTLMTFYDTGYT 526

RESULT 9

AAB97362 ID ABB97362 standard; Protein; 724 AA.

XX AC ABB97362;

XX DT 27-JUN-2002 (first entry)

XX DE Novel human protein SEQ ID NO: 630.

XX Human; antihaemic; vulnary; antiinflammatory; immunomodulator;
XX antiinfectivity; cerebroprotective; cytosolic; rheumatic; gene therapy;
XX neuroprotective; antiparkinsonian; protein therapy; EST;
XX expressed sequence tag.

OS Homo sapiens.
XX WO200222660-A2.
XX 21-MAR-2002.
XX 10-SEP-2001; 2001WO-US26015.
XX 11-SEP-2000; 2000US-0659671.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Weinman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX DR N-PSDB; AAB32548.
XX PT An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis -
XX Example 2; SEQ ID NO 630; 509pp; English.

CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate the immune system e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention.

XX SQ Sequence 724 AA;

Query Match 64.9%; Score 1254.5; DB 23; Length 724;
Best Local Similarity 48.8%; Pred. No. 7.9e-125; Indels 273; Gaps 2;
Matches 260; Conservative 0; Mismatches 0;

QY 1 EEDARSAGVATFVLQEFPRYSGYWCPKAEETPSGKILRIIYENDESEVEIIHVTSP 60
DB 186 EEDARSAGVATFVLQEFPRYSGYWCPKAEETPSGKILRIIYENDESEVEIIHVTSP 245
QY 61 METRRADSFRRYKGTANPKVTFKMSIIMDAEGRII----- 98
DB 246 METRRADSFRRYKGTANPKVTFKMSIIMDAEGRIIIVDKELIQPEILFEGVEYIA 305
QY 99 ----- 98
DB 306 RAGWTPBCKYAMSILDRSQTRLQIVLISPFLIPVEDDYMERQRLIESVPSVTPLIY 365
QY 99 ----- 98
DB 366 EETTDIWINIHDFHVPQSHBEIEFIFASECKTGRRLHYKTTSLIKESKYRSSGGLP 425
QY 99 -----VDEVRRLVYFEGTKDSPLEHLLVYVSYVN 127
DB 426 APSDFKCPIKEEIAITSGEMEVILGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLVYVSYVN 485
QY 128 PGCVTRLTDRGVSHSCCISQHCDFPISKYSNQNPKCVSLYKLSPEDDPTCKTKERWAT 187
DB 486 PGCVTRLTDRGVSHSCCISQHCDFPISKYSNQNPKCVSLYKLSPEDDPTCKTKERWAT 545
QY 188 ILDSAGPLDPYTPPELIFSFESTTGFTLYGMLYKPHLOPKKYPVLFYIGSPQVQLVNN 247
DB 546 ILDSAGPLDPYTPPELIFSFESTTGFTLYGMLYKPHLOPKKYPVLFYIGSPQVQLVNN 599
QY 248 RFGKGVYFRNLNTLASLIGYVVVVDNRGSGHRLKFGAFKYYKMGQIEIDDQVEGLQYLAS 307
DB 600 ----- 599
QY 308 RYDFIDLDVGIHGMVSYGYLSLMMQSRDIFRVAIAGAPVTLMTFYDTGYT 360

Db 600 -----|||-----VAIAGAPVTLMTIFYDTGYT 618

RESULT 10
ABB97361

ID ABB97361 standard; Protein; 782 AA.

AC ABB97361;

DT 27-JUN-2002 (first entry)

DE Novel human protein SEQ ID NO: 629.

XX Human; antihaemic; vulnerary; antiinflammatory; immunomodulator;

KM antifertility; cerebroprotective; cytosclastic; rheumatic; gene therapy;

KW neuroprotective; antiparkinsonian; protein therapy; EST;

XX expressed sequence tag.

OS Homo sapiens.

PN WO200222660-A2.

PD 21-MAR-2002.

PF 10-SEP-2001; 2001WO-US26015.

PR 11-SEP-2000; 2000US-0659671.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

DR WPI, 2002-292408/33.

XX N-PSDB; ABB97361.

PT An isolated polynucleotide for treating diseases associated with its

XX encoded polypeptide such as cancer and multiple sclerosis -

PS Example 2; SEQ ID NO 629; 509bp; English.

XX The present invention provides the protein and coding sequences of 444

CC novel human proteins. These were isolated from expressed sequences tags

CC (ESTs). They can be used to stimulate cell growth, to regulate

CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth

CC e.g. in burn treatment, to regulate the immune system e.g. to treat

CC multiple sclerosis, to regulate activin or inhibin e.g. to treat

CC infertility, to regulate haemostasis or thrombolysis e.g. to treat

CC stroke and cancer, to screen for drugs, to treat inflammatory conditions

CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.

CC Parkinson's disease. The present sequence is a protein of the invention.

XX Sequence 782 AA;

Query Match 64.9%; Score 1254.5; DB 23; Length 782;

Best Local Similarity 48.8%; Pred. No. 8.9e-125;

Matches 260; Conservative 0; Mismatches 0; Indels 273; Gaps 2;

Db 424 ERTTDIWINHIDIFHVFPQSHBEIEFIPASECKTGFRHLKYITSLKESKYRSSGGLP 483

QY 99 -----VDEVRLLYFEETKOSPLEHHLYVNSYVN 127

Db 484 APSDFKCPKIEBIATSGEMEWLGRHGSNIQVDEVRLLYFEETKOSPLEHHLYVNSYVN 543

QY 128 PGEVTRLTDGRVSHSCCISOHCDPFIISKYSNQNPKCVSLYKLSPEDDPTCKEPMAT 187

Db 544 PGEVTRLTDGRVSHSCCISOHCDPFIISKYSNQNPKCVSLYKLSPEDDPTCKEPMAT 603

QY 188 ILDSAGPLPDYTPPEIIFSESTTGFTLYGMLYXPHDLPQKKYPTVLFIYGQVQLVNN 247

Db 604 ILDSAGPLPDYTPPEIIFSESTTGFTLYGMLYXPHDLPQKKYPTVLFIYGQVQLVNN 657

QY 248 RFKGVYKFRLLNLASLGYYVVVVDNKGSCHRGLKFGCAFYKXNGQLEIDQVEGLQYLAS 307

Db 658 -----VAIAGAPVTLMTIFYDTGYT 676

QY 308 RYDFIDLDVRVGHGWSYGYLSIMLMQRSDIFRVAIAGAPVTLMTIFYDTGYT 360

Db 658 -----VAIAGAPVTLMTIFYDTGYT 676

RESULT 11

ID AAB24171 standard; Protein; 830 AA.

XX AAB24171;

DT 23-SEP-2002 (first entry)

XX Human dipeptidyl peptidase 4 (DPP4)-like 2 protein.

XX Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;

KM autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;

KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;

XX antiviral; enzyme; DPP-4 like 2 protein.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Msc-difference 235 /note= "Encoded by GAG"

PN WO200234900-A1.

PD 02-MAY-2002.

XX 29-OCT-2001; 2001WO-AU01388.

XX 27-OCT-2000; 2000AU-0001078.

XX (UNSY) UNITV STDNEY.

XX Abbott CA, Gorrell MD;

XX WPI; 2002-454646/48.

XX N-PSDB; AAB38957.

PT New dipeptidyl peptidase (DPP) peptidases, useful for screening

XX inhibitors of DPP catalytic activity, which may be employed to treat

XX e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft

XX rejection and HIV infection -

XX Disclosure; Page 82-86; 91pp; English.

XX The present invention relates to dipeptidyl peptidase (DPP) proteins and

XX polynucleotides encoding such proteins. The DPP peptidases are useful for

XX screening inhibitors of DPP catalytic activity. The inhibitors are useful

XX for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft

SQ Sequence 830 AA;

Query Match	Score	DB	Length
Best Local Similarity	63.1%	1220.5	830
Matches 236; Conservative	44.2%	Pred. No. 4.3e-121	
Mismatches 54; Indels 175; Gaps 2			

QY	2	EDARSAGVATFVLQGEFEDYSGVMWCPKAEFTTPSG- KILALILHENDSEVELIHTSP	60
Db	201	DIKSGVATFVLQGEFDRFTGTWMCPLTSMWGSQGLTKLILYEEVDESEVEIHWSP	260
QY	61	MLETRPADSFRRYPKTGTANPKVTFKMSIIMIDAEERII-	98
Db	261	ALFEERKUDSYRPRGTSSKNPKIALKLABFQTDSDQCKIVSTQEKELVQPFSSLFPKXBYIA	320
QY	99	-----	98
Db	321	RAGMTDGGKAWAMFLDRPQWLQVLPLPALFIPSTENEQRILASARAFAFRVQPYVY	380
QY	99	-----	98
Db	381	EEVTNVMVNIHFYPPQSEGEDELCLFLANECKTGFCHLYKVTALVKSQGYMSPFS	440
QY	99	-----VDEVRLLVTPBGTGKSPLEBHLVYVSIV	126
Db	441	PGDEBCKPIKEIALTSGEMEWLARHSGKIWMNEETGLVYFGQTKDPLBHHLYVVSYE	500
QY	127	NPEEVTRLTDRGYSHSCCISOCHDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMA	186
Db	501	AAGEIIVRLTTPGSHSCSMQNFDMFVSHYSVSFPCVNHYYKLSGDDDDPLKHQPRFMA	560
QY	187	TILDSAGPLDPTTPPPIFSBSTGFTLYGMLYKPHILOPBGKPYTLVLIYGGBOVOLN	246
Db	561	SMMEAAASCPPDVVPEIPIFHFTRSVRLYKMIYKPHALOPBGKPYTLVYVGGBOVOLN	620
QY	247	NREKGYKYERLNTLTASLGYVVVVIDNRGSCHRGKLFEGAFYKNGGOLEIDOVEGAOYLA	306
Db	621	NSFKGIKYLRLNTLTASLGYAAVVVIDGRSSCQRGARFBGALKNMQGOVIEIDQYBGLOFVA	680
QY	307	SRIDPFDLDRVGIGHGMSGYGLSTLMALMQSDIRRVATAGAPVTLMIFYDTGYT	360
Db	681	EKGKGFIDLSRVAIHGMSYGGFSLMGLIHKQVFRKVIAAGAPVTVMMAVYDTGYT	734

RESULT 12

ABG61592
ID ABG61592 standard; Protein; 863 AA.

AC ABG61592;

DT 12-AUG-2002 (first entry)

Human DPPIV related serine protease DPP-2.

KM Hepatitis, serum protease; dipeptidyl peptidase IV-related protein, DPR
KM DPP1; infection; human immunodeficiency virus; HIV-1, HIV-2, pain
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyskinesia; reproductive disorder; inflammatory disorder;
KM metabolic disorder.

OS Homo sapiens.

PN WO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

PR 12-OCT-2000; 2000US-240117P.

PA (FERR) FERRING BV.

XX

PI Qi S, Akinsanya KO, Riviere PJ, Junien J;

DR WPI; 2002-444178/47.
DR N-PSDB; ABK83323.

DR N-PSDB; ABK83323

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT

PS Claim 17; Fig 1; 113pp; English.

CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPPP proteins.

SQ Sequence 863 AA;

Query Match	63.1%	Score 1220.5	DB 23	Length 863
Best Local Similarity	44.2%	Pred. No. 4.6e-121		
Matches 236	Conservative	54	Mismatches 69	Indels 175
				Gaps 2

Qy	2	EDASAGCATVYLQEFPRISGYWQCPAETTPSG-KILRLIYEENDESEVNI	IHTSP	60
Db	234	DDPSAGATVILQEEFDRFTGYWQCPASWEGSGGLKRLRLIYEEVDESEVNIHVPSP	233	
Qy	61	MLETRRADSFRRPKGTANPKVTFKXSEIMIDAEGRII	-----	98
Db	294	ALERKTDTSYRRPTGSINPKIALKLAEFQTDSSGKIVSTOEKELQPPSSLPKVEYIA	353	
Qy	99	-----	-----	98

Qy 307 SRYDFIDLDRVGTHGWSYGYLSLALMQRSDIFRVALGAPVTLMIFYDTGYT 360
: ||||| : ||||| : : : ||||| : |||||
Db 714 EKKGFIDLSRVALHGWSYGYFLSMLGLHKKPQVFVAIAGAPVTWMAAYDPGYT 767

RESULT 13

ABB98134
ID ABB98134 standard; Protein; 892 AA.

Thu Oct 16 09:59:38 2003

us-10-070-464-7.rag

Page 12

[illegible]

Search completed: October 15, 2003, 17:10:48
Job time : 23.9534 secs

Job time : 23.9534 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:10:56 ; Search time 8.03173 Seconds
(without alignments)
1896.467 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1933
Sequence: 1 EEDRRSGVATFVLQEEFDR.....RVAIAGAPVTLMIFYDTGYT 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pep:*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB.pep:*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB.pep:*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB.pep:*
- 5: /cgn2_6/ptodata/1/iaa/PTCUTS COMB.pep:*
- 6: /cgn2_6/ptodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086.5	56.2	310	US-09-794-236-4	Sequence 4, Appli
2	347.5	18.0	771	US-09-462-284-2	Sequence 2, Appli
3	315	16.3	755	PCT-US93-07923-3	Sequence 2, Appli
4	315	16.3	759	PCT-US93-07923-2	Sequence 2, Appli
5	315	16.3	766	US-08-230-491A-3	Sequence 3, Appli
6	315	16.3	766	US-08-619-480A-3	Sequence 3, Appli
7	315	16.3	766	US-08-940-391-3	Sequence 3, Appli
8	315	16.3	766	US-09-794-236-1	Sequence 1, Appli
9	315	16.3	766	US-10-002-593-6	Sequence 6, Appli
10	257	13.3	760	US-08-230-491A-2	Sequence 2, Appli
11	257	13.3	760	US-08-619-280A-2	Sequence 2, Appli
12	257	13.3	760	US-08-940-391-2	Sequence 2, Appli
13	204	10.6	632	US-09-016-080-1	Sequence 1, Appli
14	196	10.1	657	US-09-355-166-1	Sequence 1, Appli
15	166.5	8.6	622	US-08-664-646A-2	Sequence 2, Appli
16	166.5	8.6	622	US-09-066-385-2	Sequence 2, Appli
17	166.5	8.6	622	US-09-261-006-2	Sequence 2, Appli
18	166.5	8.6	622	US-08-951-088-2	Sequence 2, Appli
19	166.5	8.6	622	US-09-609-566-2	Sequence 2, Appli
20	166.5	8.6	622	US-09-609-570-2	Sequence 2, Appli
21	166.5	8.6	622	US-09-427-372-2	Sequence 2, Appli
22	166.5	8.6	622	US-09-693-554-2	Sequence 2, Appli
23	148.5	7.7	593	PCT-US93-07923-11	Sequence 11, Appli
24	147.5	7.6	721	US-09-390-234-20	Sequence 20, Appli
25	131.5	6.8	344	US-09-724-623-79	Sequence 79, Appli
26	116	6.0	614	US-09-252-991A-20060	Sequence 20060, A
27	110.5	5.7	836	US-09-491-356C-21	Sequence 21, Appli

28	105	5.4	529	4	US-09-252-991A-24711	Sequence 24711, A
29	104.5	5.4	489	1	US-08-589-893-12	Sequence 12, Appli
30	104.5	5.4	489	2	US-09-062-991-12	Sequence 12, Appli
31	104.5	5.4	489	2	US-09-062-991-12	Sequence 12, Appli
32	104.5	5.4	489	2	US-09-062-991-12	Sequence 12, Appli
33	104.5	5.4	489	2	US-09-062-991-12	Sequence 12, Appli
34	102.5	5.3	489	4	US-08-124-674-2	Sequence 181, App
35	102.5	5.3	686	3	US-09-368-169-8	Sequence 2, Appli
36	101.5	5.3	298	4	US-09-355-166-6	Sequence 8, Appli
37	101	5.2	318	4	US-09-355-166-6	Sequence 16, Appli
38	98.5	5.1	444	1	US-08-483-140-28	Sequence 28, Appli
39	98.5	5.1	444	2	US-08-485-938A-32	Sequence 32, Appli
40	98.5	5.1	489	1	US-08-589-893-2	Sequence 2, Appli
41	98.5	5.1	489	1	US-08-589-893-4	Sequence 4, Appli
42	98.5	5.1	489	1	US-08-589-893-6	Sequence 6, Appli
43	98.5	5.1	489	1	US-08-589-893-8	Sequence 8, Appli
44	98.5	5.1	489	1	US-08-589-893-14	Sequence 14, Appli
45	98.5	5.1	489	1	US-08-589-893-16	Sequence 16, Appli

ALIGNMENTS

```
RESULT 1
US-09-794-236-4
; Sequence 4, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/794,236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-236-4

Query Match 56.2%; Score 1086.5; DB 4; Length 310;
Best Local Similarity 80.6%; Pred. No. 2.2e-106;
Matches 204; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY 108 FEGRKSPLEHMLVYVSNPGEVTRLTDRGYSHSCICSHCHDFPISKYSNQKNPHCVSL 167
D 1 FEGRKSPLEHMLVYVSNPGEVTRLTDRGYSHSCICSHCHDFPISKYSNQKNPHCVSL 60
QY 168 YKLSPPDDPTCKTKEFMATILDSAGPLPYTPPEIRFSFSTGTGFTLYGMLYKPHDIOPG 227
D 61 YKLSPPDDPTCKTKEFMATILDSAGPLPYTPPEIRFSFSTGTGFTLYGMLYKPHDIOPG 120
QY 228 KKTPTVLFYGGPOVLVNNRFGVKYFRNLTLASLGYVVVVDINRSGRGLKFEAGAFK 287
D 121 KKTPTVLFYGGPO----- 134
QY 288 YKKGQIFIDQVBLQYLARYPFIDLDKRGVHGMVSGYGLSLMALMQRSDIFVVAIAGA 347
D 135 --GQIEIDQVBLQYLARYPFIDLDKRGVHGMVSGYGLSLMALMQRSDIFVVAIAGA 191
QY 348 PVTLMIFYDTGYT 360
D 192 PVTLMIFYDTGYT 204

RESULT 2
US-09-462-284-2
; Sequence 2, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
```

APPLICANT: Nestec S.A.
APPLICANT: Monod, Michel
APPLICANT: Doumas, Agnes
APPLICANT: Affolter, Michael
APPLICANT: Van Den Broek, Peter
TITLE OF INVENTION: CLONING OF THE
TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
TITLE OF INVENTION: ASPERGILLUS ORYZAE
FILE REFERENCE: 8265-298
CURRENT APPLICATION NUMBER: US/09/462,284
CURRENT FILING DATE: 2000-01-03
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 771
TYPE: FRT
ORGANISM: Fungus
US-09-462-284-2

Query Match 18.0%; Score 347.5; DB 4; Length 771;

Best Local Similarity 25.4%; Pred. No. 1.1e-27; Matches 124; Conservative 57; Mismatches 150; Indels 157; Gaps 16;

QY 8 GVATFVLQEEF--DRYSGYWWCPKAEFTTSGKILRIIYENDESEVEIHHVTSPLMELTR 65
DB 190 GVPDMVYEEELIGDRYA-LWFSPOE-----YLAYLSENETGVPTTYQYVMDNQE 239
QY 66 RADSF-----RYPKGTANPKVT-----FKMSEIMT----- 91
DB 240 IAPAYPMELKIRYPVSQTNPTVTLSLNLASKVEKAPIDAFESTDLIICEVAMLDPTH 299
QY 92 -----DAEGR----- 96
DB 300 TTVAAKAKENRYDOOKVAVDPTASNKATVISDRDGTGMDNLMSKTYIGPKSDKDAY 359
QY 97 -----IIVDEVRRLVYEGTGSPLEHHL 120
DB 360 YIDISDHSGMAHLYLPVSGGEPIPLTKGDWEVTSILSDQERQVLYVLSQHHSTERHL 419
QY 121 YVSVYVNGEYTRILDRGYS--HSCCTISQHCDFPISKSNQKNPHCVSLYKLSPEDDPT 178
DB 420 YSVSY-STFPAVTPVDVTVAAYWSASFANSYVILTYGGDPVY-DELYTTNS----- 471
QY 179 CKTEFMATILDSAGPL---PDYTPPEIFSEF--STGFTLYGMLYKPHDLQPGKKYPTV 233
DB 472 --TKPL-RTITDNKAVLEQIDYALPNTTYFELPLPSGETLNMQRPLPGSSPDKKYFIL 528
QY 234 LFIYGGPOVQLVNNRPFKGVKY-PRINTLASLGYVVVVIDNRGSGHRLKFEGARFYKXGQ 292
DB 529 FTFYGGPQAQCEVTRKQWALNFKAYVASDSELEYVTWVDNRGTGFGKGRKFRSAVTRQLGL 588
QY 293 IEIDDVQVGLQYLASRYFIDLDRVGHGMSYGYLSMALMORSDFRVALAGAPVTLW 352
DB 589 LEAEQYLAAQO-QANIPWIDADHIGLWMSFGGLISKVLEKSGAFTLVITRAPVSDW 647
QY 353 IFYDTGYT 360
DB 648 RFYDSWYT 655

RESULT 3

PCT-US93-07923-3

Sequence 3, Application PC/TUS9307923

GENERAL INFORMATION:

APPLICANT: Morimoto, Chikao

APPLICANT: Schlossman, Stuart F.

APPLICANT: Tanaka, Toshiaki

TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESS: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07923
FILING DATE: 19930819
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,162
FILING DATE: 21-AUG-1992
APPLICATION NUMBER: 07/832,211
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 755
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07923-3

Query Match 16.3%; Score 315; DB 5; Length 755;

Best Local Similarity 23.4%; Pred. No. 3e-24; Matches 120; Conservative 59; Mismatches 146; Indels 188; Gaps 22;

QY 1 EEDARSQVATFVLQEE-FDRYSGYWWCPKAEFTTSGKILRIIYENDESEVEIHHVTSPLMELTR 56
DB 179 KEDIYNGITDWEVEEFVSAYSLAWSPNGTF-----LVAQNDTEVPLIEYSF 229
QY 59 VTSPLMELTRRADSRFPYKGTANPKVTF-----KMSIMIDAEGRITIDEVVR 104
DB 230 YSDSLQPKYKTVRYPYKAGAVNPTVAFVNVNTDLSSTVNMATSIQITAPRSMILG----- 285
QY 105 LVYFEGTKDSPELHLLVVSIVNPGVET---RLTRGYSHSC-----CI--SQ 147
DB 286 -----DHYLDDVTWATQERISLQWLRRIQYVSVMDICDYDESSGRWNCIVARQ 333
QY 148 HCDF---FISKYSNQNPHCV---SLYKLSPEDE-----DPTCKTKFMA 186
DB 334 HIEMSTGWGRF-RPEEPHTLDGNSFYKILSNBEGYRHICYFOIDKXCTPITKTWE 392
QY 187 TI----- 188
DB 393 VIGIEALTSPLYIYSNEYKMPGGRMLYKQLSDYTKVTCLSCELNPERCOYYSVSFSK 452
QY 189 -----IDSAGP-LPDYT-----PEIISF--ESTTG 211
DB 453 EAKYQLRCSGPGPLVTLHSSVNDKGLRVLEDNSALDKMLQNVQMPBKLDFTILLNETK 512
QY 212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPOVQLVNNRPFKGVYFRLN---TLASLGYVVV 268
DB 513 F-WYQMLPRH-FKSKKYPLILLDVYAGPCQ-----KADTVRLNMATYLASTENIIV 564
QY 269 V-IDNRGSGHRLKFEGARFYKXGQIIBDQVEGLQYLASRYFIDLDRVGHGMSYGY 327
DB 565 ASFDGRSGGYQGDKIMAIRRLQTFEVEDQIEAARQF-SMGVVDNRKRIATWMSYGY 623
QY 328 LSLMALMORSDFRVALAGAPVTLIMIFYDGYT 360
DB 624 VTSNVLGSGGVFRGGLAVAPVSRWEYDSVYT 656

```

Query Match      16.3% Score 315; DB 5; Length 759;
Best Local Similarity 23.4%; Pred. No. 3e-24;
Matches 120; Conservative 59; Mismatches 146; Indels 188; Gaps 22.

QY      1 EEDARSAGVATFVLQOE-PRDRYSGYMCPKAETTPSGGKILRIITYEENDESEVEIIT-- 56
Db      183 KEDIIYNGIITDWMYEEVEFSAYSALMWSPNGTF-----LAAQGNDDREVLPIEVSF 233
QY      57 VTSFMLETRRDSFRYPKGTANKRYTF-----KMSEIMDAEGRIYDEVRR 104
Db      234 YSDESLQPKTKVRYVPAGAVNLPVKFFVYVNTDSLSSVTNATSIQITAPASMLG---- 289
QY      105 LVVEEGTKDSPLEHLLVYVSYNQGEV----RLTRDGYGSHC-----CI--SQ 147
Db      290 -----DHYLCGVTMATGERISLTQMLRIQNSVWDICDYDESSGRMNCTVARQ 337
QY      148 HCDF-----FISKYGNQKNPHCV---SIKYLSPEPD-----DPTCKTKEFWA 186
Db      338 HIEMSTTGWGRF-RPSEPHFTLDGNSFYKLIISNEBGRHICYPQIDKKDCTFITKGTWE 396
QY      187 TI----- 188
Db      397 VIGHEALTSPLYIYISNEYKMGPGRNLIYKIQLSDYTKVTLSCSEINPERQIYVSRSK 456
QY      189 -----LDSAGP-LPDYIT-----DPEIISF--ESTTG 211

```

```

1      US-08-230-491A-3
2      RESULT 5
3      Sequence 3, Application US/08230491A
4      Patent No. 5587299
5      GENERAL INFORMATION:
6      APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
7      APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
8      TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
9      TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
10     NUMBER OF SEQUENCES: 10
11     CORRESPONDENCE ADDRESSES:
12     ADDRESSEE: FELSE & LYNCH
13     STREET: 805 THIRD AVENUE
14     CITY: NEW YORK
15     STATE: NEW YORK
16     COUNTRY: USA
17     ZIP: 10022
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
20     COMPUTER: IBM PS/2
21     OPERATING SYSTEM: PC-DOS
22     SOFTWARE: WORDPERFECT - ASC II
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/08/230,491A
25     FILING DATE: 20-APRIL-1994
26     CLASSIFICATION: 435
27     ATTORNEY/AGENT INFORMATION:
28     NAME: Hanson, No. 5587299man D.
29     REGISTRATION NUMBER: 30,946
30     REFERENCE/DOCKET NUMBER: LUD 330
31     TELECOMMUNICATION INFORMATION:
32     TELEPHONE: (212) 688-9200
33     TELEFAX: (412) 838-3884
34     INFORMATION FOR SEQ ID NO: 3:
35     SEQUENCE CHARACTERISTICS:
36     LENGTH: 766 amino acids
37     TYPE: amino acid
38     TOPOLOGY: linear
39     IS-08-230-491A-3

```

	Query Match	16.3%	Score 315	DB 1	Length 766	
	Best Local Similarity	23.4%	Pred. No. 3.1e-24			
	Matches 120	Conservative 59	Mismatches 146	Indels 188	Gaps 22	
QY	1	BEDARSAQAVATVLQEE-FDRYSGYWCQPAETTPSGKILRIIYEENDESEVEIITH----	56			
DB	190	KEDIIYNGITDWMYEEEFASYSALMWSPNQTF-----LAYAQFNDTEVPLEIYSF	240			
QY	57	VTSPLLETRRRDSEFRYPKRTGIANPKYTF-----KNSELMIDAGRIIYDEVRR	104			
DB	241	YSDSESLQYPKVAVRVPYFAGAVNPTVYFVWNTDLSISVYNNATSIQITAPASKLIG----	296			
QY	105	IYVEFGTKDSEPLEHLLVYASVYNGEYTF-----RLDRGRSHSC-----CI--SQ	147			
DB	297	-----DHFLCDVTWATQOERISLOWMLRIQYISVMDICDYDESSGRNCCIVARQ	344			
QY	148	HCDP-----FLISKYSNOKNPHCV-----SLYKLSPED-----DPTCKTKEFA	186			


```

; INFORMATION FOR SEQ ID NO: 3
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 766 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;
US-08-940-391-3

```

Query Match	16.3%;	Score 315;	DB 2;	Length 766;
Best Local Similarity	23.4%;	Pred. No. 3.1e-24;		
Matches 120; Conservative	59;	Mismatches 146;	Indels 188;	Gaps 22;

[illegible]

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000


```

; CURRENT FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-016-080-1

```

```

Query Match      10.6%; Score 204; DB 3; Length 632;
Best Local Similarity 26.3%; Pred. No. 1.2e-12;
Matches 83; Conservative 53; Mismatches 123; Indels 56; Gaps 17;

```

```

QY      67 ADSFRY-PRGTANPKVTEKM-----SEIMDAEGRIVDEVARLVEFG-----TKDSP 115
      250 AHVHYDPKTKELK-KLTKDLDMNAYNSINSDVRG-----SQRAELVYKEGMVIYVATDGP 304
QY      116 LEHLVYVSNVNGEVTRLT--DRGSHSCCISQCHDFITSKSNQKNHCYSLYKLSPP 173
      305 -RANLFRVNL--DGKLERVYGGDRSVE-----SPDIGLIAFTADAVPTLTY 352
QY      174 EDDPTCKTEF--WATILDSAGLPDYT--PREIFSFESTTGFTLYGMLYKPHLDQPKK 229
      353 RQKEKKTVDENKM-----IKGYTLKPEHFVKASDGEIDAWMKPVNPKGKK 403
QY      230 YPLVLFITYGPOVQVNNRFGKYKF-RNLTLASLGVYVVDNRSSCHRGKLFEGAFY 288
      404 YPALLETIHGGPKTAY-----GYAFMEHFVLTLSKGFVIFSNRSGSDGGEFF-ADIRG 456
QY      289 KKGQIIRIDQVEGLQYLAIRYDPIIDRYGIGHWSYGGVLSLMAKQSRDIFPVATA GAP 348
      457 HYGERPYQIMEVVDALRRFPIDGRILGVTGSGYGGFWTM-IVGHTNRFAAVTQNS 515
QY      349 VTLWIFY----DTGY 359
      516 ISNWISFGFTIDIGY 530
DB

```

```

RESULT 14
US-09-355-166-1
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolyase-Fold Enzymes
; FILE REFERENCE: GCS11-PCT
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-09-355-166-1

```

```

Query Match      10.1%; Score 196; DB 4; Length 657;
Best Local Similarity 24.9%; Pred. No. 9.1e-12;
Matches 69; Conservative 43; Mismatches 107; Indels 58; Gaps 9;

```

```

QY      111 TDSPLLEHLVYVSNVNGEVTRLTDRG-----YSHSCCISQCH 149
      311 TDS---QGFYVIG-----TDQSTGIYIISIEGLVPIRLKRYINSFSLSPDE 357
DB      150 DEFFISKYSNQKNPHCVSLYKLSPPEDPTCKTEFMATILDSAGLPDYT--PREIFSE 207
      358 QHFISAVTKPDRPSSELSIPLQGEKEQLTGANKF-----VREHTISPERIQYA 407
QY      208 STTGFTLYGMLYKPHLDQPKKYPTVLFYGGPOVQVNNRFGKYKF-RNLTLASLGV 266
      408 TEDGVVNVGNWLMKRPAMQEGETTYPLILNIHGGPHMMY-----GHTYFHEFOVLAKKGYA 461
DB

```

```

QY      267 VVVIDNRGSGCHRGKLFEGAFKRYKQIETIDQVEGLQYLAIRYDPIIDRYGIGHWSYGG 326
      462 VVYINFRSGHGYGQEFVNAVREDYGGKDDVQWQAVDEAIKRPDHPRLGVTGSGYGG 521
QY      327 YLSMALMQRSDIFRVAIAGAPVTLWIFY----DTGY 359
      522 FMTNM-IVGQINRFAKAYIQRSISNWISHGVSIDIGY 557
DB

```

```

RESULT 15
US-08-664-646A-2
; Sequence 2, Application US/08664646A
; Patent No. 5877001
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amidases
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARILLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/664,646A
; FILING DATE: June 17, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Heron
; REGISTRATION NUMBER: 28, 019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-664-646A-2

```

```

Query Match      8.6%; Score 166.5; DB 2; Length 622;
Best Local Similarity 29.6%; Pred. No. 1.1e-08;
Matches 48; Conservative 33; Mismatches 68; Indels 13; Gaps 6;

```

```

QY      201 PPIFSFESTGTGFTLYGMLYKPHLDQPKKYPTVLFYGGPOVQVNNRFGKYKYPRLNTL 260
      352 PPHFRFKS-KDLKIGWYLRPEVKE--EKAPVIVFHGGPK-GMYGHRF---VYEMQIM 403
DB      261 ASLGVVYVVIDNRGSGCHRGKLFEGAFKRYKQIETIDQVEGLQYLAIRYDPIIDLRVGTH 320
      404 ASKGYVYVFNFRGSDGSDPALKVLEFTGTLEDFEDINNGIEEPFKLEPQADRERVGIT 463
QY      321 GWSYGYLSLIMQMQRSDIFRVAIAGAPVTLWIFY----FYDTG 358
      464 GISTGGFMTNMALTLQ-SDLFKAGISENGISYVLTGYAFSDIG 504
DB

```

```

Search completed: October 15, 2003, 17:15:40
Job time : 10.0317 secs

```

QY 99 ----- 98
Db 424 EETTDIMINIHDI FHVFPQSHHEEIEFIFASECKTGFRHLYKITSILKESKYRSSGGLP 483
QY 99 ----- VDEVRLVYFEGTKOSPLEHHLYVVSYN 127
Db 484 APSDFKCPIKEIAITSGEMEWLGRHOSNIQVDEVRRLVFBGTKOSPLEHHLYVVSYN 543
QY 128 PGEVTRLDRGYSHSCCISQHCDFPISKYSNQNKPCHVSLYKLSPPEDDPTCKTKFPMAT 187
Db 544 PGEVTRLDRGYSHSCCISQHCDFPISKYSNQNKPCHVSLYKLSPPEDDPTCKTKFPMAT 603
QY 188 ILDSAGPLDPTPEPEIFSFESTTGFTLYGMLYKPHLOPKKYPVTLFIYGGPQVOLLVN 247
Db 604 ILDSAGPLDPTPEPEIFSFESTTGFTLYGMLYKPHLOPKKYPVTLFIYGGPQVOLLVN 663
QY 248 RFGVXYFRNLNTASLGYYVWVVDNRGSCHRGLKFBGAFYKXGQIEIDQVGLQYLAS 307
Db 664 RFGVXYFRNLNTASLGYYVWVVDNRGSCHRGLKFBGAFYKXGQIEIDQVGLQYLAS 723
QY 308 RYDFIDLRVGHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYT 360
Db 724 RYDFIDLRVGHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYT 776

RESULT 2
US-10-054-776-2
; Sequence 2, Application US/10054776
; Publication No. US20030165818A1
; GENERAL INFORMATION:
; APPLICANT: Mark Robert Edbrooke
; APPLICANT: Alan Peter Lewis
; TITLE OF INVENTION: NOVEL PROTEIN
; FILE REFERENCE: Q6104205
; CURRENT APPLICATION NUMBER: US/10/054, 776
; NUMBER OF SEQ. ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ. ID NO 2
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-776-2

Query Match 95.0%; Score 1836.5; DB 12; Length 882;
Best Local Similarity 67.5%; Pred. No. 5,7e-179;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
QY 1 EEDARSAGVATFVLOEBFDRYSGVWMCCKAETTPSGKILRIIYENDESEVELIHVTS 60
Db 244 EEDARSAGVATFVLOEBFDRYSGVWMCCKAETTPSGKILRIIYENDESEVELIHVTS 303
QY 61 MLETRRADSRYPKVTGTANPKVTPKXSEIMIDAGRLL----- 98
Db 304 MLETRRADSRYPKVTGTANPKVTPKXSEIMIDAGRLLIIVDKELIQPFIILFEGVEYIA 363
QY 99 ----- 98
Db 364 RAGWTBEGKYAMSIILDRSQTRLQIVLISPELFI PVEDDVERORLIESVDSVTPLLIY 423
QY 99 ----- 98
Db 424 EETTDIMINIHDI FHVFPQSHHEEIEFIFASECKTGFRHLYKITSILKESKYRSSGGLP 483
QY 99 ----- VDEVRLVYFEGTKOSPLEHHLYVVSYN 127
Db 484 APSDFKCPIKEIAITSGEMEWLGRHOSNIQVDEVRRLVFBGTKOSPLEHHLYVVSYN 543
QY 128 PGEVTRLDRGYSHSCCISQHCDFPISKYSNQNKPCHVSLYKLSPPEDDPTCKTKFPMAT 187
Db 544 PGEVTRLDRGYSHSCCISQHCDFPISKYSNQNKPCHVSLYKLSPPEDDPTCKTKFPMAT 603

QY 188 ILDSAGPLDPTPEPEIFSFESTTGFTLYGMLYKPHLOPKKYPVTLFIYGGPQVOLLVN 247
Db 604 ILDSAGPLDPTPEPEIFSFESTTGFTLYGMLYKPHLOPKKYPVTLFIYGGPQVOLLVN 663
QY 248 RFGVXYFRNLNTASLGYYVWVVDNRGSCHRGLKFBGAFYKXGQIEIDQVGLQYLAS 307
Db 664 RFGVXYFRNLNTASLGYYVWVVDNRGSCHRGLKFBGAFYKXGQIEIDQVGLQYLAS 723
QY 308 RYDFIDLRVGHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYT 360
Db 724 RYDFIDLRVGHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYT 776

RESULT 3
US-10-170-789-38
; Sequence 38, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Welch, Nadine
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
; FILE REFERENCE: 10448-191001
; CURRENT APPLICATION NUMBER: US/10/170, 789
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/797, 039
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06525
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186, 061
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 09/882, 166
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19269
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212, 078
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/934, 406
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26052
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226, 740
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/861, 801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16549
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205, 508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/801, 267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187, 454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829, 671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197, 508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961, 721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235, 023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045, 367
; PRIOR FILING DATE: 2001-11-07

```
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-789-38
```

```
Query Match      95.0%; Score 1836.5; DB 12; Length 882;
Best Local Similarity 67.5%; Pred. No. 5,7e-179;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
```

```
QY 1 EEDASAGVATFVLQEEFDRYSGYMWCPKATTPSGKILRIIYEENDESEVEIHHVTSF 60
Db 244 EEDASAGVATFVLQEEFDRYSGYMWCPKATTPSGKILRIIYEENDESEVEIHHVTSF 303
QY 61 MLETRADSFRRYKGTANPKVTFKXSEIMDAEGRII----- 98
Db 304 MLETRADSFRRYKGTANPKVTFKXSEIMDAEGRII----- 98
QY 99 ----- 98
Db 364 RAGWTEGKYAMSLIDRSOTRQIYLISPELFIPEDDVMERQRLIESVPSVTPLIIT 423
QY 99 ----- 98
Db 424 EETTDIMINIDIFHVFPQSHHEIEFIIPASECKTFRHLKYITSLIKESKYRSGGLP 483
QY 99 ----- 98
Db 484 ASDDFKCP1KEBIAITSGEMEWLGRHGSNIQVDEVRLVYFSGTKOSP1EHLHYVSYN 543
QY 128 PSEVTRLTDRGYSHSCCISOHCDPFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWAT 187
Db 544 PSEVTRLTDRGYSHSCCISOHCDPFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWAT 603
QY 188 IIDSAGPLPDYTPPEIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQVQVLNN 247
Db 604 IIDSAGPLPDYTPPEIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQVQVLNN 663
QY 248 RFKGVYFRPLNTLASLGYVWVVDNRGSGCHRLKFEKGFYKMGQIEIDDOVEGLOYLAS 307
Db 664 RFKGVYFRPLNTLASLGYVWVVDNRGSGCHRLKFEKGFYKMGQIEIDDOVEGLOYLAS 723
QY 308 RYDFIDLDVRVGIHGMVSYGYLSMALMORSDFRVAIAGAPVTLMIFYDTGYT 360
Db 724 RYDFIDLDVRVGIHGMVSYGYLSMALMORSDFRVAIAGAPVTLMIFYDTGYT 776
```

```
RESULT 4
US-09-976-674-3
; Sequence 3, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEIY
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 863
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-3
```

```
Query Match      63.1%; Score 1220.5; DB 10; Length 863;
Best Local Similarity 44.2%; Pred. No. 7,7e-116;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;
```

```
QY 2 EEDASAGVATFVLQEEFDRYSGYMWCPKATTPSGKILRIIYEENDESEVEIHHVTSF 60
Db 234 DDEKASGAVATFVLQEEFDRYSGYMWCPKATTPSGKILRIIYEENDESEVEIHHVTSF 293
QY 61 MLETRADSFRRYKGTANPKVTFKXSEIMDAEGRII----- 98
Db 294 ALERKTDRIYRPTKSNPKIALKLAEFQTSQKATVSQEKELVQPFSSLPKRYVIA 353
QY 99 ----- 98
Db 354 RAGWTEGKYAMSLIDRSOTRQIYLISPELFIPEDDVMERQRLIESVPSVTPLIIT 413
QY 99 ----- 98
Db 414 EEVYTNVWVNDIFVFPQSHHEIEFIIPASECKTFRHLKYITSLIKESKYRSGGLP 473
QY 99 ----- 98
Db 474 PSEVTRLTDRGYSHSCCISOHCDPFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWA 533
QY 127 NPGEVTRLTDRGYSHSCCISOHCDPFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWA 186
Db 534 AAGEIVRLTTPGFSSHSCSMQNDMFVSHSVSTPCVHVYLSGEPDPLKQPRFWA 593
QY 187 TILDSAGPLPDYTPPEIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQVQVLNN 246
Db 594 SMWEASCPDPDYTPPEIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPQVQVLNN 653
QY 247 NPFGVYFRPLNTLASLGYVWVVDNRGSGCHRLKFEKGFYKMGQIEIDDOVEGLOYLAS 306
Db 654 NPFGVYFRPLNTLASLGYVWVVDNRGSGCHRLKFEKGFYKMGQIEIDDOVEGLOYLAS 713
QY 307 RYDFIDLDVRVGIHGMVSYGYLSMALMORSDFRVAIAGAPVTLMIFYDTGYT 360
Db 714 EKYGFIDLSRVALIHGMVSYGYLSMALMORSDFRVAIAGAPVTLMIFYDTGYT 767
```

```
RESULT 5
US-09-976-674-23
; Sequence 23, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEIY
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-23
```

```
Query Match      63.1%; Score 1220.5; DB 10; Length 892;
Best Local Similarity 44.2%; Pred. No. 8,1e-116;
```

Matches	236;	Conservative	54;	Mismatches	69;	Indels	175;	Gaps	2
QY	2	EDARAGVATVLOEFPDRYSGYWMCPKAEITPSSG-KILRLIYEENDESEVEIITHVSP	:	:	:	:	:	:	60
Db	263	DDPKSAGATVIOEFPDRFTGYWMCPTASWEGSGELKTLRLIYEVEDESEVEIIVHSP	:	:	:	:	:	:	322
QY	61	MLETRASSFRPKFKGTANPKVTFPMSEIMDABERIT	:	:	:	:	:	:	98
Db	323	ALBEKRTDSYRPRGSKNPKIALKLAIEFQDTSQCKIVSTQEKELVDPFSSLPFKVEYIA	:	:	:	:	:	:	382
QY	99	-----	:	:	:	:	:	:	98
Db	383	RAGWTRDCKVAMAMFLDRPQOWLVLPLPALFIISTENEBQRLASARAVRNQPYVY	:	:	:	:	:	:	442
QY	99	-----	:	:	:	:	:	:	98
Db	443	EEVTNVMINVDIEFYPQSEGEDELCEFLRANECKTGCHIKYTAVLKSGQYDWSSEPS	:	:	:	:	:	:	502
QY	99	-----	:	:	:	:	:	:	98
Db	503	PGDEFFKPIKEELITGTGEMVLRASHSKITWNETKLVFGKDPLEHHLVVSYE	:	:	:	:	:	:	562
QY	127	NPEVETRLTDRGYSHSCISQCHDFEISKYSNQKNPHCVSLYKXSSPEDDPCKTKEFWA	:	:	:	:	:	:	186
Db	563	AAGEIVRLITTEGFSSCMSQNFDFVSHYSSVSYPCHVHYKLSGPDPPPLHKQPRFWA	:	:	:	:	:	:	622
QY	187	TIIDSAGLPLYTPPEIIFESFSTGTFLYGMLYKXHDLOPGKKYPTVLFIYGGPQVLVN	:	:	:	:	:	:	246
Db	623	SMMEASCPDPVPEIIFHFTSRDVRILYGMLYKXHALQPGKAPTLFVYGGPQVLVN	:	:	:	:	:	:	682
QY	247	NRFKGVKRYRLNTIATSLGYVVVVVIDNRSGCHRGLEFEGAFYKXKMQIIEIDQVEBLQYIA	:	:	:	:	:	:	306
Db	683	NSFKGIKYLRLNTIATSLGYAVVVIDGRGSCQGRLEFEGALKNQMGQVEIENQVEBLQYVA	:	:	:	:	:	:	742
QY	307	SRYPFIDLRGCIHWSGYGHLISLMALMQRSDIFPVYATAGAPVTLMTIYDTGYT	:	:	:	:	:	:	360
Db	743	EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKQVFFVYALAGAPVYVMAAYDTGYT	:	:	:	:	:	:	796
RESULT 6									
US-09-976-674-27									
Sequence 27, Application US/09976674									
Patent No. US20020115843A1									
GENERAL INFORMATION:									
APPLICANT: Qi, Steve									
APPLICANT: Akimsanya, Karen									
APPLICANT: Riviere, Pierre									
APPLICANT: Junier, Jean-Louis									
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV									
FILE REFERENCE: 70669									
CURRENT APPLICATION NUMBER: US/09/976,674									
CURRENT FILING DATE: 2001-10-12									
PRIOR APPLICATION NUMBER: US 60/240,117									
PRIOR FILING DATE: 2000-10-12									
NUMBER OF SEQ ID NOS: 61									
SOFTWARE: PatentIn version 3.1									
SEQ ID NO 27									
LENGTH: 892									
TYPE: PRT									
ORGANISM: Homo sapiens									
US-09-976-674-27									
Query Match 63.1%; Score 1220.5; DB 10; Length 892;									
Best Local Similarity 44.2%; Pred. No. 8.1e-116;									

[illegible]

QY 110 -----GKDSPLHHLVYVSVNPGSVRLTDRCY 139
DB 503 PGDEKCPKIEIALITSGMEVILARRSGKXTDPLEHHLVYVSYAAEIVRLTTPGF 562
QY 140 SHSCCISQCHDFPISKYNOKNPHCVSLYKLSPEDEPTCKTEFWATILDSAGPLPDYT 139
DB 563 SHSCMSQNDMVMVSHYSSVSTPCVHYVYLKSGDDPLHKQPRFMAISMEALSCPPDYV 622
QY 200 PEIIFSESTTGTFLYGMLYKPHDLPQKKYPTVLYFGSPQVQVLVNNRFKGVYFRLNT 259
DB 623 PEIIFHHTRSQVRLVGMIVKPHALPGKXHPVLYFYGGSPQVQVLVNNRSKGIKYLANT 682
QY 260 LASLGYVAVVINDRGSCCHRLKFEAGAFKYMGGIETDDQVEGLQYLASRYDFDLDVRGI 319
DB 683 LASLGYAVVVIDRGSCQRLRFEAGALKQMGQVEIEDQVEGLQFVAEKYGFIDLRSVAI 742
QY 320 HGMSYGGFLSLMALMQRSDIFRYAIAAGAPVTLMIIFYDTGYT 360
DB 743 HGMSYGGFLSLMGLIHKQVFKVAILAGAPVTLVMAVDTGYT 783

RESULT 8

US-09-976-674-35
; Sequence 35, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-35

Query Match 62.1%; Score 1201; DB 10; Length 879;
Best Local Similarity 44.3%; Pred. No. 7.9e-114;
Matches 224; Conservative 53; Mismatches 72; Indels 162; Gaps 3;

QY 2 EDARSAGVATFVLQEEFDRYSGYWMCCKAETTPSGG-KILRIIYENDESVEIITHVTS 60
DB 263 DDEKSAGVATFVLQEEFDRYSGYWMCPTASWEGSEGLKILRIIYEEVDESEVEIHHVPS 322
QY 61 MLETRRADSPRYPKTGTPANPKVTFKMSIIMIDAEGRILIVDEVRLL----- 105
DB 323 ALBERKTDSRYRTGSKNPKIALKLAEPQDSQKIVSQEKELVOPRSSLPKXVEYIA 382
QY 106 -----VY 107
DB 383 RAGWTRDGKYAMAFIDRPOWLQVLVLPALFIPSTENBQRLASRAVBRVAVQPVVY 442
QY 108 FE----- 109
DB 443 EEVTVNWINVHDIFFPPQSEGEDELCTLRANCKTGFCRLYKVTAVLKSQGYDWSEPPS 502
QY 110 -----GKDSPLHHLVYVSVNPGSVRLTDRCY 139
DB 503 PGDEKCPKIEIALITSGMEVILARRSGKXTDPLEHHLVYVSYAAEIVRLTTPGF 562
QY 140 SHSCCISQCHDFPISKYNOKNPHCVSLYKLSPEDEPTCKTEFWATILDSAGPLPDYT 199
DB 563 SHSCMSQNDMVMVSHYSSVSTPCVHYVYLKSGDDPLHKQPRFMAISMEALSCPPDYV 622
QY 200 PEIIFSESTTGTFLYGMLYKPHDLPQKKYPTVLYFGSPQVQVLVNNRFKGVYFRLNT 259

DB 623 PEIIFHHTRSQVRLVGMIVKPHALPGKXHPVLYFYGGSPQVQVLVNNRSKGIKYLANT 682
QY 260 LASLGYVAVVINDRGSCCHRLKFEAGAFKYMGGIETDDQVEGLQYLASRYDFDLDVRGI 319
DB 683 LASLGYAVVVIDRGSCQRLRFEAGALKQMGQVEIEDQVEGLQFVAEKYGFIDLRSVAI 742
QY 320 HGMSYGGFLSLMALMQRSDIFRYAIAAGAPVTLMIIFYDTGYT 360
DB 743 HGMSYGGFLSLMGLIHKQVFKVAILAGAPVTLVMAVDTGYT 783

RESULT 9

US-09-976-674-19
; Sequence 19, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-19

Query Match 62.1%; Score 1199.5; DB 10; Length 658;
Best Local Similarity 58.0%; Pred. No. 7.3e-114;
Matches 239; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

QY 1 EEDARSAGVATFVLQEEFDRYSGYWMCCKAETTPSGGKILRIIYENDESVEIITHVTS 60
DB 244 EEDARSAGVATFVLQEEFDRYSGYWMCCKAETTPSGGKILRIIYENDESVEIITHVTS 303
QY 61 MLETRRADSPRYPKTGTPANPKVTFKMSIIMIDAEGRILIVDEVRLL----- 98
DB 304 MLETRRADSPRYPKTGTPANPKVTFKMSIIMIDAEGRILIVDEVRLL----- 363
QY 99 ----- 98
DB 364 RAGWTRDGKYAMAFIDRPOWLQVLVLPALFIPSTENBQRLASRAVBRVAVQPVVY 423
QY 99 ----- 98
DB 424 BETTIDWINIHDIFFVFPQSEHEIEIFIRASECKTGFRHLKYKTSILKESKXRSGLP 483
QY 99 -----VDEVRLLVYFEGTKDSPLEHHLVYVSVN 127
DB 484 APSEDFKCPKIEIALITSGMEVILARRSGNINQVDEVARLVYFEBTKOSPLEHHLVYVSVN 543
QY 128 PGVTVRLTDGRYSHSCCISQCHDFPISKYNOKNPHCVSLYKLSPEDEPTCKTEFWAT 187
DB 544 PGVTVRLTDGRYSHSCCISQCHDFPISKYNOKNPHCVSLYKLSPEDEPTCKTEFWAT 603
QY 188 IIDSAGPLPDYTPPEIIFSESTTGTFLYGMLYKPHDLPQKKYPTVLYFYGG 239
DB 604 IIDSAGPLPDYTPPEIIFSESTTGTFLYGMLYKPHDLPQKKYPTVLYFYGG 655

RESULT 10

US-09-976-674-11
; Sequence 11, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:

```
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-11
```

```
Query Match      62.1%; Score 1199.5; DB 10; Length 661;
Best Local Similarity 58.0%; Pred. No. 7.4e-114;
Matches 239; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
```

```
QY      1 EDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEIIHTSP 60
        |||
Db      244 EDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEIIHTSP 303
        |||

QY      61 MLETRRADSRFRYPKTGTANPKVTFKXSEIMIDAGRII----- 98
        |||
Db      304 MLETRRADSRFRYPKTGTANPKVTFKXSEIMIDAGRIIIVDKELIQFELLFPGVEYIA 363
        |||

QY      99 ----- 98
Db      364 RAGWTPGKYAMSILLDRSQRLQIVLISPFLIPVEDDVMERQRLIESVPDSVTPLIY 423
        |||

QY      99 ----- 98
Db      424 EETTDWINIHDIHFVFPQSHHEIEFIPIASECKTGRRHLKYKTSILKESKYKRSGLP 483
        |||

QY      99 -----VDEVRLVYFEGTKDSPLEHHLVVSYN 127
        |||
Db      484 APSDFKCPIKEIAITSGEWEVLGRHGSNIQVDEVRRLVFEGRKDSPLEHHLVVSYN 543
        |||

QY      128 PGEVTRLTDGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPPEDDPCTKKEFWAT 187
        |||
Db      544 PGEVTRLTDGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPPEDDPCTKKEFWAT 603
        |||

QY      188 ILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDIQPGKPYTVLFIYGG 239
        |||
Db      604 ILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDIQPGKPYTVLFIYGG 655
        |||
```

```
RESULT 11
US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-7
```

```
Query Match      62.1%; Score 1199.5; DB 10; Length 690;
Best Local Similarity 58.0%; Pred. No. 7.9e-114;
Matches 239; Conservative 0; Mismatches 0; Indels 173; Gaps 1;
```

```
QY      1 EDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEIIHTSP 60
        |||
Db      244 EDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEIIHTSP 303
        |||

QY      61 MLETRRADSRFRYPKTGTANPKVTFKXSEIMIDAGRII----- 98
        |||
Db      304 MLETRRADSRFRYPKTGTANPKVTFKXSEIMIDAGRIIIVDKELIQFELLFPGVEYIA 363
        |||

QY      99 ----- 98
Db      364 RAGWTPGKYAMSILLDRSQRLQIVLISPFLIPVEDDVMERQRLIESVPDSVTPLIY 423
        |||

QY      99 ----- 98
Db      424 EETTDWINIHDIHFVFPQSHHEIEFIPIASECKTGRRHLKYKTSILKESKYKRSGLP 483
        |||

QY      99 -----VDEVRLVYFEGTKDSPLEHHLVVSYN 127
        |||
Db      484 APSDFKCPIKEIAITSGEWEVLGRHGSNIQVDEVRRLVFEGRKDSPLEHHLVVSYN 543
        |||

QY      128 PGEVTRLTDGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPPEDDPCTKKEFWAT 187
        |||
Db      544 PGEVTRLTDGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPPEDDPCTKKEFWAT 603
        |||

QY      188 ILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDIQPGKPYTVLFIYGG 239
        |||
Db      604 ILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDIQPGKPYTVLFIYGG 655
        |||
```

```
RESULT 12
US-09-976-674-29
; Sequence 29, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-29

Query Match      58.6%; Score 1132.5; DB 10; Length 832;
Best Local Similarity 42.3%; Pred. No. 7.7e-107;
Matches 222; Conservative 52; Mismatches 76; Indels 175; Gaps 2;
```

QY 99 ----- 98
Db 443 EEWYTNWVNHDIFFPPPOSEGEDELCLFLRANECKTGFCMLKYTAVLKSQGYDMSBPFS 502
QY 99 ----- 126
Db 503 PGDEDFPCPIKEEIALTSGEMEWLARHGSKIWNEEKLYVFOGTQKOTPLEHHLVYVSYE 562
QY 127 NPGVTRLTDRGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPEDEDDPTCKTEFMA 186
Db 563 AAGEIVRLTTPGFHSHSCSMQNFDMFVSHYSVSTPCVHYKLSGDDDDPLHKQPRFMA 622
QY 187 TILDSAGPLDPYTPPELFFSESTTGFTLYGMLYKPHDLOPKKPTVLYFTYGGPOVOLVN 246
Db 623 SMMEAACPPDYPPELFFHHTRSADVRLYGMKYKPHALOQPKKHPVLYFYGGPOVOLVN 682
QY 247 NRPFGVKYFRILNTLASIGYVVVYIDNRGSCHRGLKFEAGAFYKMGQIEIDDOVEGLQYLA 306
Db 683 NSRFKGYKRLNTLASIGYVVVYIDNRGSCORGLRFGALKNQMGQVIEIDQVEGLQFVA 742
QY 307 SRYPFIDLDVRVGHGWSYGGYLSMALMORSDFRVALAGAPVTL 351
Db 743 EKYGFIDLSRVALHGWMSGYGLSLMGLIHKQVKAQPLAVPRL 787

RESULT 13
US-09-976-674-31
; Sequence 31, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-31

Query Match 58.6%; Score 1132.5; DB 10; Length 832;
Best Local Similarity 42.3%; Pred. No. 7,7e-107;
Matches 222; Conservative 52; Mismatches 76; Indels 175; Gaps 2;
QY 2 EDARSAGVATFVLQEEFDRYSGYWCPCKAETTPSGG-KILRIIYEENDESEVEIITHVTS 60
Db 263 DDEKSAGVATFVLQEEFDRFTGYWCPTASWEGSEGLKILRIIYEENDESEVEIIVHPS 322
QY 61 MLETRRADSPRYPTGTANPKVTFRKSEIMIDAGRIT----- 98
Db 323 ALERKKTDSYRYPRTGSKNPKIALKLAFOQDSQKIVSQEKLVOPFSSLPFKVEYIA 382
QY 99 ----- 98
Db 383 RAGNTRDGKXAMAMFLDRPOQWQLVLLPALFIPSTENEBQRLASARAVERNOPYVYV 442
QY 99 ----- 98
Db 443 EEWYTNWVNHDIFFPPPOSEGEDELCLFLRANECKTGFCMLKYTAVLKSQGYDMSBPFS 502
QY 99 ----- 126
Db 503 PGDEDFPCPIKEEIALTSGEMEWLARHGSKIWNEEKLYVFOGTQKOTPLEHHLVYVSYE 562
QY 127 NPGVTRLTDRGYSHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPEDEDDPTCKTEFMA 186

Db 563 AAGEIVRLTTPGFHSHSCSMQNFDMFVSHYSVSTPCVHYKLSGDDDDPLHKQPRFMA 622
QY 187 TILDSAGPLDPYTPPELFFSESTTGFTLYGMLYKPHDLOPKKPTVLYFTYGGPOVOLVN 246
Db 623 SMMEAACPPDYPPELFFHHTRSADVRLYGMKYKPHALOQPKKHPVLYFYGGPOVOLVN 682
QY 247 NRPFGVKYFRILNTLASIGYVVVYIDNRGSCHRGLKFEAGAFYKMGQIEIDDOVEGLQYLA 306
Db 683 NSRFKGYKRLNTLASIGYVVVYIDNRGSCORGLRFGALKNQMGQVIEIDQVEGLQFVA 742
QY 307 SRYPFIDLDVRVGHGWSYGGYLSMALMORSDFRVALAGAPVTL 351
Db 743 EKYGFIDLSRVALHGWMSGYGLSLMGLIHKQVKAQPLAVPRL 787

RESULT 14
US-09-976-674-37
; Sequence 37, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-37

Query Match 57.6%; Score 1113; DB 10; Length 819;
Best Local Similarity 43.0%; Pred. No. 7,5e-105;
Matches 220; Conservative 51; Mismatches 79; Indels 162; Gaps 3;
QY 2 EDARSAGVATFVLQEEFDRYSGYWCPCKAETTPSGG-KILRIIYEENDESEVEIITHVTS 60
Db 263 DDEKSAGVATFVLQEEFDRFTGYWCPTASWEGSEGLKILRIIYEENDESEVEIIVHPS 322
QY 61 MLETRRADSPRYPTGTANPKVTFRKSEIMIDAGRITVDEVR----- 105
Db 323 ALERKKTDSYRYPRTGSKNPKIALKLAFOQDSQKIVSQEKLVOPFSSLPFKVEYIA 382
QY 106 ----- 107
Db 383 RAGNTRDGKXAMAMFLDRPOQWQLVLLPALFIPSTENEBQRLASARAVERNOPYVYV 442
QY 108 PE----- 109
Db 443 EEWYTNWVNHDIFFPPPOSEGEDELCLFLRANECKTGFCMLKYTAVLKSQGYDMSBPFS 502
QY 130 ----- 139
Db 503 PGDEDFPCPIKEEIALTSGEMEWLARHGSKIWNEEKLYVFOGTQKOTPLEHHLVYVSYE 562
QY 140 SHSCCISQCHDFPISKYSNOKNPHCVSLYKLSPEDEDDPTCKTEFMA1ILDSAGPLDPYT 199
Db 563 SHSCSMQNFDMFVSHYSVSTPCVHYKLSGDDDDPLHKQPRFMA5MMEAACPPDYV 622
QY 200 PPELFFSESTTGFTLYGMLYKPHDLOPKKPTVLYFTYGGPOVOLVNNRFGVKYFRILNT 259
Db 623 PPELFFHHTRSADVRLYGMKYKPHALOQPKKHPVLYFYGGPOVOLVNNRFGVKYFRILNT 682
QY 260 LASIGYVVVYIDNRGSCHRGLKFEAGAFYKMGQIEIDDOVEGLQYLASRYFIDLDRCI 319

D6 683 IASLGTAIVVIVIDRGSCQQRGLRFEGLAKKQMGQVETIEDQVEGLQFVAEKIGFIDLSKVAI 742

D5 320 HGMSYGGYISLMAIMQSDIPRVAATAGAVTL 351
|||||:||||: : : :
743 HGMSYGGFISLMGLIHKPVQFKAQPLAPPRLL 774 .

RESULT 15

```

US-09-976-674-39
/ Sequence 39, Application US/09976674
/ Patent No. US20020115843A1
/ GENERAL INFORMATION:
/ APPLICANT: Q1, Steve
/ APPLICANT: Akinsanya, Karen
/ APPLICANT: Riviere, Pierre
/ APPLICANT: Junien, Jean-Louis
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLIV
/ FILE REFERENCE: 70669
/ CURRENT APPLICATION NUMBER: US/09/976,674
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,117
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 39
/ LENGTH: 819
/ TYPE: FRT
/ ORGANISM: Homo sapiens
/ US-09-976-674-39

```

Query Match	57.6%;	Score 1113;	DB 10;	Length 819;
Best Local Similarity	43.0%;	Pred. No. 7.5e-105;		
Matches 220;	Conservative 51;	Mismatches 79;	Indels 162;	Gaps 3;

QY	2	EDRARSGLVATPVYLOEJFEDSYSGWMCCKAETTPSSG-KILSLVEENDESVEIIVHTSP	60
Dd	263	DDKSKAGVATPVYLOEJFEDFRTGYWMCCTPASMESGSLKTLILILBEYDESEVEIVHPS	3222
QY	61	MIETRADSFYRYPKGTGTANPKVTFKMSIIMDAEGRIIVDEVRIL-----	1055
Dd	323	ALBERTDTSYRYPRITSGKNPKITALKLAEFQTSQCKIVSTQEKELVDPFSSLPEKVEYIA	3822
QY	106	-----VY 107	
Dd	383	RAGWTEGKYAMAFILDRPOQWLQVLLPPLFLPSTENBQRLASAPAVRNQYVVY	4422
QY	108	FE-----109	
Dd	443	EEVTNVMVNIHFYRPQSEGEDELCPLANECKTGFCHLYKTVAVLKSGYDMSBFS	5022
QY	110	-----GKSDPLFHHLLVYVSYYNPGVTRLTLTRGY	139
Dd	503	PGDEKCPRIKEBIALTSGEMEWLARHSGSKTKPDLPHLLHYVSYEARAGIYALLTPPGF	5622
QY	140	SHSCCISOHCFPIISKYSNOKNPKCHVSLYKLSPEDDPDTCKTEKFEMATILDSAGPLDPT	199
Dd	563	SHSCMSQNDPMVSHYSSVSTRPCVHYKLSGRDDPLHNPQRFMSMDEAASCPEDYV	6222
QY	200	PEPIISFESESTGTLTYGMLYKPHDLOPKKPYTYVLFYSGRQVOLLVNNRFGVKYFFPLNT	259
Dd	623	PEPIIFFHTRSDVRLTYGMLYKPHALOPEKKEPITYLVYGGROYOLLVNNSEFGIIXYTLRNT	682
QY	260	LASLGVVVVVVINDRGSCRGRLFEFGAFKYMQLIRIDQVAGLOYLASRYFDILDRVGI	319
Dd	683	LASLGAVVVVVINDRGSCQGRGLAFBKALKKNQMQVAILIEQVZGLQFVAEKYGFIDLSVALI	742
QY	320	HGMSYGGYISLMAIMQSRDIEFVVAIAGAVTL 351	
Dd	743	HGMSYGGYISLIMGLIHKPQVFAQPLAVPRL 774	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:08:55 ; Search time 9.45959 Seconds

(without alignments)
3659.853 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1933

Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVALAGAPVTLMIFYDTGYT 360

Scoring table: BLOSUM62

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	527.5	27.3	738	2	A87516 dipeptidyl peptida
2	491.5	25.4	931	2	T32919 hypothetrical prote
3	430.5	22.3	741	2	JC5142 X-Pro dipeptidyl-p
4	369	19.1	711	2	S66261 X-Pro dipeptidyl-p
5	315	16.3	766	1	CDH26 dipeptidyl peptida
6	312.5	16.0	793	2	T41703 dipeptidyl aminope
7	309	16.0	760	1	S23752 dipeptidyl-peptida
8	291.5	15.1	792	1	A39914 dipeptidyl-peptida
9	280.5	14.5	818	1	A30107 dipeptidyl aminope
10	250	12.9	931	2	A49737 dipeptidyl aminope
11	245	12.7	795	2	F82658 dipeptidyl-peptida
12	240	12.4	829	2	T19514 hypothetrical prote
13	226.5	11.7	683	2	E87495 prolyl oligopeptid
14	221	11.4	803	2	A41793 dipeptidyl aminope
15	220	11.4	709	2	B82580 alanyl dipeptidyl
16	219	11.3	779	2	T25173 hypothetrical prote
17	219	11.3	799	2	T25174 hypothetrical prote
18	205.5	10.6	631	2	H75007 probable acylamino
19	204	10.6	642	2	C71337 hypothetrical prote
20	203	10.5	803	2	T68600 dipeptidyl aminope
21	203	10.5	865	2	T54331 dipeptidyl aminope
22	199	10.3	622	2	F71174 hypothetrical prote
23	197.5	10.2	743	2	T37700 probable dipeptidyl
24	196	10.1	657	2	E70025 probable dipeptidyl
25	193	9.8	632	2	E75057 peptidase PAB1418
26	189	9.0	536	2	F90299 acylaminoacyl-pept
27	174	9.0	759	2	I38593 acylaminoacyl-pept
28	166.5	8.6	591	2	B84381 acylaminoacyl-pept
29	165	8.5	591	2	H72474 probable acylamino

30	162.5	8.4	569	2	S74053 probable acylamino
31	149	7.7	676	2	C97775 acylamino-acid-rel
32	147.5	7.6	667	2	A87711 prolyl oligopeptid
33	147.5	7.6	721	2	T09631 probable acylamino
34	140.5	7.3	591	2	A99444 acylaminoacyl-pept
35	140.5	7.3	606	2	T35378 probable peptida
36	139.5	7.2	745	2	T33751 hypothetrical prote
37	138.5	7.2	598	2	F84199 hypothetrical prote
38	137.5	7.1	654	2	AD3183 peptidase (impor
39	134.5	7.0	732	1	JC4655 acylaminoacyl-pept
40	131.5	6.8	732	1	S07624 acylaminoacyl-pept
41	130.5	6.8	572	2	F72455 probable acylamino
42	129.5	6.7	555	2	AD1794 acylase and dieste
43	127.5	6.6	555	2	AB1419 acylase and dieste
44	126.5	6.5	629	2	T15945 hypothetrical prote
45	126.5	6.5	689	1	JN0585 prolyl oligopeptid

ALIGNMENTS

RESULT 1
A87516
dipeptidyl peptidase IV [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: A87516
R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg
N.J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: A87516
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-738 <STO>
A/Cross-references: GB:AE005673; NID:G13423647; PIDN:AAK24125.1; GSPDB:GN00148
C/Genetics: 0
A/Gene: CC2154

Query Match	27.3%	Score 527.5;	DB 2;	Length 738;
Best Local Similarity	28.5%;	Pred. No. 1e-36;		
Matches 137;	Conservative 61;	Mismatches 132;	Indels 151;	Gaps 13;
QY	2	EEDARSAGVATFVLQEEFDRYSGYWMCFAETTSQGIILRLVYENDESEVEIIHVTSPM 61		
DB	189	KDALSRGVAEFIVQEEIDRFTGYWSPDES-----RIYTRVDESGVDIV----- 233		
QY	62	LETRRAD-----SFRYPKTGTANP-----KVTF----- 84		
DB	234	---PRADIGGATVVGQRYPGRAPRAVVDLFFVRDLASGVTALDIGNANDIYVARVAM 290		
QY	85	-----KME-----IMD----- 92		
DB	291	SADGKTIVYVQRLSRDOKTLDLAFDAATGAKXTLLTDTPHFIEVSNDFRPLDTGTFLMG 350		
QY	93	-----AEGRIL-----VDEVRLLVYEGTKDSPLHNHLYVS 124		
DB	351	SEKDGNHLRYRAADGKLIQITKGDMPVGLGECVBARVAFSAITDPIERRLYVS 410		
QY	125	YNNPGEVATRLDRGSHSCISQHCDFEISKYSNOKNPHCVSLKSSP-----EDDPTC 179		
DB	411	YARPKPKALTSAGGWWAAKVAANGGAFACTYSDPKTPSQTALYSADGKRVRIENKLA 470		
QY	180	KTEFPNATLDSAGPLPDYTPREIFPESFTGTLYGMLYKPHDLQGGKPYTLFTYGG 239		
DB	471	EGHPYW-----PYAANPQ-----PEFGSLKADGSLTHLILKPIGFDPAAKYPALVSYVG 523		
QY	240	PQVQLVNNRFGKVKYFRLNTLASLGGVVVVVINDRSGCHRGLEKGAFFKYKGGIETIDOV 299		
DB	524	PHAQRVKQMKWHSPE---RTYLEAGVVIKFLDNDRSGNSAKFMRLDRKLTGYVEVDQL 580		

A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1711 <RAB>
 A:Cross-references: EMBL:D42121; NID:9577283; P1DN:BA07702.1; PID:9577284
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidylpeptide hydrolase

Query Match 19.1%; Score 369; DB 2; Length 711;
 Best Local Similarity 23.4%; Pred. No. 3,1e-23;
 Matches 112; Conservative 70; Mismatches 147; Indels 142; Gaps 13;

```

QY      8 GVAFTVLQEEFDYRSYGYWCPKAEFTTPSGKILRLILEENDESEVEI-IHVT-S--PMLET 64
      172 GLDGMVYEEFGHADDYQW-----NKAQDALVVRDEKVPKPEINIPYQNIYPLMT 225
      65 RRADSPRYPKTGANPKVT-----FKNS-----IMDAGRIIVDEV 102
      226 ----YKYPKAGENSAYTAVLYQLSGSKSAQLNFGSSKKYVLPQLFQTNANDEIVATA 280
      103 RR-----LVYFGTKDSPLE-----HLLY- 121
      281 NRHONKVDLKVNTKTAASVKLFTETDNAMLETDNLTMEPLDINSFLMASERDGHRLTYW 340
      122 ----VVS-----VNPGEVTLTD 136
      341 YDAAGKLKQVSGKDWELINYYGVNPKTEVYIQTEKKSINKVVSKLNTINTGKTQLSN 400
      137 KGVSHSCCICQCHDFPISKSNQKNPCHVSLYKLSPE-----DDPCKTKEFWATIL 189
      401 AEGNNSAASFSTKNTYFINTSSTAKVPKTYLKDKANGKDVVELQNNDDLKLKLS----- 454
      190 DSAGPLPDYTPPELFFESFTTGFTLYGMLKYPKPHLDQGKKYPTVLFYGGPQVQVNNR 249
      455 -----DNFLAKFFITIPNAGQOMAMMTKPKNFPDAKRYPYFMFYSGPSQOVANSW 508
      250 KGVYFRLNLTASLGYYVIVIDNRSGCHGLKEFGAPKYMGQLEIDQVEGLQYASRY 309
      509 DGGNGIWMFLMLAQKGLVVCVDGRGTGFRGTGKTKKVTYKRLGKKEIEDQITAKMLGNQ- 567
      QY      310 DPFIDLVRVGHGWSYGYSLMALMQRSDIFRVALAGAPVTLMIFFDTGYT 360
      568 SYVDKSRIGIFGWSYGGYVNASLMTKGDVFKMGIAVAPVTNRFYSIYT 618
      DB
  
```

RESULT 5

CDH26 dipeptidyl-peptidase IV (EC 3.4.14.5) - human

N:Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence revision 23-Aug-1996 #text change 18-Jun-1999
 C:Accession: S24313; B42408; B61136; S59510; I56154; S59857; S15520

R:Mitsumi, Y.; Hayashi, Y.; Arkawa, F.; Ikehara, Y.
 Biochim. Biophys. Acta 1131, 333-336, 1992
 A:Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a ser

A:Reference number: S24313; PMID:92339551; PMID:1352704
 A:Accession: S24313

A:Molecule type: mRNA

A:Residues: 1-6, 'I', '8-766 <MIS>

A:Cross-references: EMBL:X60708; NID:G35335; P1DN:CAA3118.1; PID:G35336

R:Barbat, A.; J. Biol. Chem. 267, 4824-4833, 1992

A:Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer

IV mRNA levels during cell differentiation.

A:Reference number: A42408; PMID:92165847; PMID:1347043

A:Accession: B42408

A:Molecule type: mRNA

A:Residues: 1-5, 'R', '7-436, 'S', '438-556, 'I', '558-662, 'E', '664-766 <DAR1>

A:Cross-references: GB:M80536; NID:g181569; PIDN:AAA52308.1; PID:g181570

A:Experimental source: intestine

A>Note: this sequence corresponds with the author's translation

A:Accession: A42408

A:Molecule type: mRNA

A:Residues: 1-5, 'R', '7-436, 'S', '438-556, 'I', '558-662, 'E', '664-711, 'G', '713-766 <DAR2>

A:Cross-references: GB:M80536; NID:g181569
 A>Note: sequence extracted from NCBI Backbone (NCBI:N:83986, NCBI:P:83988); this sequen

R:Gorvel, J.P.; Ferrero, A.; Chamraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.

Gastroenterology 101, 618-625, 1991

A:Title: Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human small

A:Reference number: A61136; PMID:91317403; PMID:1677636

A:Accession: B61136

A:Molecule type: protein

A:Residues: 1-15, 'X', '17-22 <GOR>

R:Boehm, S.K.; Gum Jr., J.R.; Erickson, R.H.; Hicke, J.W.; Kim, Y.S.

Biochem. J. 311, 835-843, 1995

A:Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from

A:Reference number: S59510; PMID:9606759; PMID:7487939

A:Accession: S59510

A:Molecule type: DNA

A:Residues: 1-31 <BOB>

A:Cross-references: GB:S79876; NID:g1195574; PIDN:AA835614.1; PID:g1195575

R:Tanaka, T.; Camerini, D.; Seed, B.; Torimoto, Y.; Dang, N.H.; Kameoka, J.; Dahlber

J. Immunol. 149, 481-486, 1992

A:Title: Cloning and functional expression of the T cell activation antigen CD26.

A:Reference number: I56154; PMID:92325476; PMID:1352530

A:Accession: I56154

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-436, 'S', '438-766 <TAN>

A:Cross-references: GB:M74777; NID:g180082; PIDN:AAA51943.1; PID:g180083

R:Abdolt, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.

Immunogenetics 40, 331-338, 1994

A:Title: Genomic organization, exact localization, and tissue expression of the human

A:Reference number: S59857; PMID:95012454; PMID:7927537

A:Accession: S59857

A:Molecule type: DNA

A:Residues: 1-436, 'S', '438-766 <ABB>

A:Cross-references: EMBL:U13734

C:Genetics:

A:Gene: GDB:DBP4

A:Cross-references: GDB:125239; OMIM:102720

A:Map position: 2q24.3-2q24.3

A:Insertions: 2/3; 32/1; 65/1; 95/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 35

C:Superfamily: dipeptidyl-peptidase IV

C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; homodimer; proteinase; transme

F:1-6/Domain: intracellular #status predicted <INT>

F:7-28/Domain: transmembrane #status predicted <TM>

F:29-76/Domain: extracellular #status predicted <EXT>

F:85,92,150,219,229,281,321,520,685/Binding site: carbohydrate (Asn) (covalent) #statu

F:630,708,740/Active site: Ser, Asp, His #status predicted

Query Match

Best Local Similarity 16.3%; Score 315; DB 1; Length 766;
 Matches 120; Conservative 59; Mismatches 146; Indels 188; Gaps 22;

```

QY      1 EEDPARAGVATFVLQEE-FDQVSGYWCPEKAEFTTPSGKILRLILEENDESEVEI-IH- 56
      190 KEDIIYNGITDWMVEEIVSAVYALWMSFNGTF-----LAKQFNTQVPLIEYSF 240
      DB
      57 VTSPLMELTRADSPRYPKTGANPKVTF-----KMEIMIDAGRIIVDEVRR 104
      DB
      241 YSDSESLQYKTVKVPYKAGAVPTVKFFVNTDLSVYTNATSIQITAPASMLIG----- 296
      QY      105 LVYFEGTKDSPLEHHIYVSVYVNGEVT-----RLNDRGSHSC-----CT--SQ 147
      DB
      297 -----DHYLCDTWTATQERISIQWLEIRIONYSVMDICDYDESSGRNCLVARQ 344
      QY      148 HCDP-----FISKYSNOKNPCHV-----SLYKLSSEP-----DPTCKTKEFWA 186
      DB
      345 HIEMSTTGWVGGR-RSEBHFPLDNGSFYKILISNEGYNHICFQJDXKDCITITIGTWE 403
      QY      187 TI----- 188
      DB
      404 VIGIEALTSDYLYVISNEYKMPGGRNLYKIQLIDYTKVTLCSCEINPERCQYYSFSK 463
      QY      189 -----LDSAGP-LPDYT-----PPEIFSF--DSTTG 211
  
```


Keywords: dipeptidylpeptidase hydrolase, glycoprotein, homodimer, liver, serine protease
F.1.792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental
F.1.728/Domain: signal sequence #link MATS #status experimental <Sig>
F.1.6/Domain: intracellular #status predicted <Int>
F.7-28/Domain: transmembrane #status predicted <TM>
F.29-792/Domain: extracellular #status predicted <EXT>
F.29-34/Domain: propeptide #link MATS #status experimental <PRO>
F.73-792/Product: dipeptidyl-peptidase, soluble form #status experimental <MATS>
F.83-90,149,217,227,319,521,686/Binding site: carboxylate (Asn) (covalent) #status predicted
F.631/Active site: Ser #status experimental
F.709,741/Active site: Asp, His #status predicted

Best Local Similarity 24.1%; Pred. No. 1,4e-16;
Matches 119; Conservative 61; Mismatches 153; Indels 161; Gaps 22;

A:Molecule type: DNA
A:Residues: 1-82, 'H', 84-124, 'N', 126-181, 'LRRL'ET', 189-199, 'N', 201-365, 'DPKRGKERKF', 376-574
A:CROSS-references: EMBL:X15484
A>Note: the authors translated the codon ACC for residue 572 as Asn
C:Genetics:
A:Gene: SGD:DAP2; STE13; MIPS:YHR028C
A:CROSS-references: SGD:S0001070; MIPS:YHR028c
A:Map position: 8R
C:Superfamily: dipeptidyl-peptidase IV
C:keywords: dipeptidyl-peptide hydrolase; glycoprotein; transmembrane protein; yeast vacuole F30-45/Dominant transmembrane #status predicted <TM>
F30-45/Dominant transmembrane #status predicted <TM>
F33-79,110,139,392,421/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.5%; Score 280.5; DB 1; Length 818;
Best Local Similarity 22.0%; Pred. No. 1.3e-15;
Matches 114; Conservative 50; Mismatches 155; Indels 199; Gaps 19;

Dy 8 GVATFLVLOEF-PDRYSGYWCPKAETTPSGKIILRIILEYNDESEV-EIHVTSPMLETR 65
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dz 234 GKPMVVVEEHEFEEDDKAAAMSPPTGD-----YLAFLKIDEEVEGFEFI--IPYYVD 281
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 66 RAD-----SRYPKTGTANP----- 80
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 282 EKDIYPEPKRSIKYEKSQTNPFAELIIVTSMDDGISPHRISGNKKDSLLTEVTWGNG 341
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 81 ---KVTFKMSEI---MID----- 92
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 342 NVLVKTTDRSSDLITFLPIDTIAKTSNVVRNESSNGWEITHNTLFIPANETFDPPHG 401
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 93 -----AEGRVIYDEVRLVPEFGTDSPLEH 118
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 402 YVDILPIPGVNHHLAVFNENSSSHVKTLTEGWENVANGPLADSENRLVFISTKSSSTER 461
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 119 HLIVSVSYNPSGEVTRLTLDGRYS--HSCCIISOHCDFIFISKYSNQKNPH----- 163
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 462 HVYIIDARSPNEIIIEVDTSDGVADVVSFGSRRFFGLITYGPRVPYOKIIVDFHSRAEK 521
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 164 C-----VSLKYSSPEDDPTCKTKERWATILLASAGLPDYTPPETIRFESSTTGFTLYG 216
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 522 CDKGNIAGLSHYHEKKN-----VLTKI--LEDYAVERFSFRFLNDGKEBFG 566
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 217 --MYKPHDLQP-----GKKYPTVLFIYGAPQOVVNNSFKVGKPYRLN----TIASL 263
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 567 KDILVNSEILPNPDFDELSDHYEVFPFYAYGSPNSQY-----VKITSVGFNEVASQL 620
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 264 GVVVVVIDNRGSCHREGLFEQAIFYKMGOEIFIDQVEGLQYLAARYFDIDLDRVGHGWS 323
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 621 NAIIVVVVGRGTFXGQDFRSILVRDLIGDYEARDDISAASYGS-LRFVDPOKISLFCWS 679
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 324 YGGYSLMALMQRSD-IFFVAIAGAPTYLMTFYDNGYT 360
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 680 YGGYLLTKLERDGGRHFHKYGMNAVPTDMRPYDSVT 717
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 10
A49737
dipeptidy] aminopeptidase [EC 3.4.14.-] - Yeast (*Saccharomyces cerevisiae*)
N:Alternative names: Protein OS045; protein YOR219c; protein YOR50-9
C:Species: Saccharomycetes cerevisiae
CjDate: 03-Mar-1994 #sequence revision 03-Mar-1994 #text change 29-Oct-1999
CjAccession: A49737; S45451; S60946; S67112; S71721
R:Santa Anna-A, S.; Herskowitz, I.
submitted to the Protein Sequence Database, July 1993
A:Reference number: A49737
A:Accession: A49737
A:Molecule type: DNA
A:Residues: 1-931 <SN>
A:CROSS-references: GB:I21944; NID:G347196; PIDN:AAA35119.1; PID:G347197
R:Ama-Arriola, S.S.; Herskowitz, I.
Yeast 10, 801-810, 1994
A>Title: Isolation and DNA sequence of the STE13 gene encoding dipeptidy] aminopeptidase
A:Reference number: S45451; MUID:95066582; PMID:7975897
A:Accession: S45451

A:Molecule type: DNA
A:Residues: 1-931 <ANN>
A:Cross-references: EMBL:L1944; NID:g347196; PIDN:AAA5119.1; PID:g347197
R:Galisson, F.; Dujon, B.
Submitted to the EMBL Data Library, October 1995
A>Description: Sequence and analysis of a 33 Kb fragment from the right arm of chromo
A:Reference number: S60938
A:Accession: S60946
A:Molecule type: DNA
A:Residues: 1-931 <GBL>
A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA6182.1; PID:g1050771
R:Boyer, J., Fairhead, C.; Gallion, L.; Galisson, F.; Michaux, G.; Thierry, A.: Dujon
Submitted to the Protein Sequence Database, July 1996
A:Reference number: S67104
A:Accession: S67112
A:Molecule type: DNA
A:Residues: 1-931 <BOY>
A:Cross-references: EMBL:Z75127; NID:g1420507; PIDN:CAA99437.1; PID:e252398; PID:g1422
R:Galisson, F.; Dujon, B.
Submitted to the EMBL Data Library, October 1995
A:Experimental source: strain S288C
A:Cross-reference: EMBL:Y92441; NID:g1050762; PIDN:CAA63182.1; PID:g1050771
A:Residues: 1-931 <GMW>
A:Cross-references: EMBL:Y92441; NID:g1050762; PIDN:CAA63182.1; PID:g1050771
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Function:
A:Description: involved in processing of alpha-factor prepropheromone
Keywords: dipeptidyl-peptide hydrolase; glycoprotein; transmembrane protein
F.II3-150/Doman: transmembrane #status predicted <TM>
F.I37/Binding site: carbohydrate (Asn) (covalent) #status predicted
F.I785, I863/Active site: Ser, Asp, His #status predicted

Query Match 12.9%; Score 250; DB 2; Length 931;
Best Local Similarity 24.3%; Pred. No. 6e-13;
Matches 101; Conservative 56; Mismatches 162; Indels 94; Gaps 15;

Oy 15 QEEFDTRYGYMCP--KAETPSGGKILRLYEENDESEVEIIHVS----- 59
Db 434 OKDSILLNGKKISPDTRFRFEITDRNSKILDVVKVDIPSSQWLTVRNNTSNLFNGMIERTK 493

Oy 60 -----PMLERR-----DSDS-----FRPKGTGANPKTFFKMSEMIDAEGRII 98
Db 494 DIISIPEPPELKRDVGDIIDIADSRGFSLFYTP-VFAKEPIQLTKGWMEVTGNIGIVG 552

Oy 99 VDEVRAIRLYVEGSTDSLHLHYLVSYV-----NPGEVTLTDRGYSHCIS 146
Db 553 YEVEDTDIFTFANIEGVMSQHLYAISLTDTQTONTPOSLONP-----SPKYDFEDEL 606

Oy 147 OHCFPFISKTSNQNP-----HCVSLYLKSSPPEDPTCKTEFWATILD 190
Db 607 SSARVAASKKLGPDPFKVAGPLRVINVAEHDDSILOLTKE----KFKE----- 654

Oy 191 SAGLPYDTPPEIPSFSST---GFTLYGMLYPKHDOPGKGKYPVLFIYGAPVOLVN 247
Db 655 ---XIKNYDLP-TISXYTMLWDGOVBELNIYEIFKPANIHPKKKIYLIUNVIIGPSSQIFTT 710

Oy 248 RFKVAYKFLNLTAASLGVVAVVIDNRSGCHRLKFECAGFYKMGQILEIDDQE-GLOYLA 306
Db 711 --KSLSIAFEQAQVASGLDVIVIQIEPRGTGKGWSFRSMAREKLGWEPREDITEVTKRFIQ 768

Oy 307 SRVFILDLVRGIHGWSYGYLLMAL-MORSIDIERVALICARPVTLMIFYDGYT 360
Db 769 RNSGHIDESKIALIMGSYGGFTSLKTVELNQGTFKYAMA VAFVMTNLTYDSVT 823

RESULT 11

F82858

dipeptidyl-peptidase XF0015 [imported] - Xylella fastidiosa (strain 9a5c)

C/Species: Xylella fastidiosa

C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C/Accession: F82858

R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing 406, 151-157, 2000

A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A/Reference number: A82515; MUID:20365717; PMID:10910347

A/Note: for a complete list of authors see reference number A59328 below

A/Accession: F82858

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-795 <SIM>

A/Cross-references: GB:AE003856; GB:AE003849; NID:99104770; PIDN:AAF82828.1; GSPDB:GN001

A/Experimental source: strain 9a5c

R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Britton, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H.

as-Neto, E.; Docena, C.; El-Dorzy, H.; Facincani, A.P.; Ferreira, A.J.S.

Submitted to GenBank, June 2000

A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurume, E.E.; Laiz

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.

Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A/Reference number: A59328

A/Contents: annotation

A/Genetics: XF0015

C/Genetics:

Query Match 12.7%; Score 245; DB 2; Length 795;

Best Local Similarity 27.5%; Pred. No. 1,3e-12;

Matches 98; Conservative 49; Mismatches 140; Indels 70; Gaps 16;

QY 21 YSGYWMCPKAEITPSGGKILRIIYENDESEVEIIHVTSPMLETRADSPRYPKGTANP 80

DB 396 FNDPFGMLPDNRIT-----LMLSEBSGSHLYTIQ-----ETGPRKQ 431

QY 81 KATP---KMSIMIDAAGRIIVDEVRRLVYFEGKSPLEHMLVYSVYVNG--EYTRL 134

DB 432 R-TFGHWEVSAPPPSADG-----RSVYF-----VCPITWPGYEVCKLD 469

QY 135 TDGYSHSCISQHC-DFISKYSNQKNPCHVSLYKLS-SPEDPTCKTKEFNATIL--- 189

DB 470 LDGSHFDTLALQGVDFDLSPDGRQ---ILVLYSSYLPRQLAVISTNGGARVITDT 525

QY 190 -DSAGLPDYTPPEISFEESTTG-FLYKMLYKPHDIQPGKRYPTVLFIYGGPOVOLVNN 247

DB 526 RTSTFRKQWIOPEYVQIPSKHGAGVIMGKYGPASIQPEKRYPIVFWFHAGAYLQVNSA 585

QY 248 RFKGVYERLNTLASL---GYVVVVDNRGSGCHRGKFGKAFYKMGQIIDDQVBLQ 303

DB 586 RYP-SIFRQMFHTLLVQGYVLDDIPASAGHGDMWTALYRNNGHPELDYLDGLD 643

QY 304 YLASRYDFIDLDEVGIGHWSYGYLSIMALMORSDFRVAIAGAPVTLMIFDYDTGYT 360

DB 644 WLVAHKGQ-DRRRAGWYGSYGSGFMTMAFRAPGTFKAAALRPVDMQVNHET 699

RESULT 12

T19514

hypothetical protein C27C12.7 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C/Accession: T19514

R/Thomas, K.

submitted to the EMBL Data Library, March 1996

A/Reference number: Z19134

A/Accession: T19514

A/Status: preliminary; translated from GB/EMBL/DBD8

A/Molecule type: DNA

A/Residues: 1-829 <WIL>

A/Cross-references: EMBL:Z69883; PIDN:CAA93743.1; GSPDB:GN00028; CESP:C27C12.7

A/Experimental source: clone C27C12

C/Accession: E87495

R/Nerman, W.C.; Feilblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K

n, U.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of Caulobacter crescentus.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: E87495

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-683 <STO>

A/Cross-references: GB:AE005673; NID:913423453; PIDN:AAK3961.1; GSPDB:GN00148

A/Genetics:

A/Genetics: CC1986

C/Genetics:

Query Match 11.7%; Score 226.5; DB 2; Length 683;

Best Local Similarity 21.1%; Pred. No. 3.7e-12;

Matches 107; Conservative 77; Mismatches 144; Indels 178; Gaps 20;

QY 8 GVATFVLOEE-FRYSQYWMCPKAEITPSGGKILRIIYENDESEVEIIHVTSPMLETRR 66

DB 249 GLANMLYEBITLSSAVVM-----SPSGRYVSYLRFDDREVNRIPLPKTID----- 296

QY 67 ADSP-----RYPKTGTANPKVTFKMSIMIDAAGRIIVDEV----- 102

DB 297 -DSYVEYFELPYPKAGVQNTL---VTQYIMDSNNHKIVETAPNELSAANGDYVLTNK 352

QY 103 -----RLVYFEGTQDSPLEHMLV----- 122

DB 353 WITPFRGSDLGERLVTWANRD---QNHVYFSLCNEODCVNALSPOFSIDNRQLWVSP 409

QY 123 --VSYPNGEVTLT-----DRG--YSH-----SCCTSGHCFPT----- 153

DB 410 KDVKGVPFETGTGLTVLPHHGDGNTYNNHAHVELDGTGKTKWIGENFDVLVLGYS 469

QY 154 -----SKYSNQKNPHCVSLYK-----LSPEDDPT 178

DB 470 SKIDALTFSAVGDVGEFSTYIVREAWYSNK-----TTLQKVTDOFEDCKTLGSGADPT 525

QY 179 -----CK-----TFEFTATLIDSA-----GPLEDYPPRLP-SPESTGTGLY 215

DB 526 GORIVVQCEKPFNTRLYLVDDVDTTKKMLBEGTKAVIPFDVNMKFGKIKLPSGIDGH 585

QY 216 GMLYKPHDLOPGKRYPTVLFIYGGPOVOLVNNRFEKGYKPRLNTLASLGYVVVVIDRGS 275

DB 586 YMLTFRNLNDGAKIPLIDITYGSPDSKQVFOKTPPTMAAIQISQYDIAY--ARIDRGT 643

QY 276 CHRGLKFEAGFKYKMGQIIDDQVEGLQYLASRYDFIDLDVRGIGHWSYGYL-SLMALM 334

DB 644 GGRGMDYKEAVYRKLGDIAEVDTLDMIRAFINFGFIDBRILAWGMSYGGFLTSKIAIK 703

QY 335 QRSDFRVAIAGAPVTLMIFDYDTGYT 360

DB 704 DQGLVKAISIAPIVTDPKYDSAYT 729

RESULT 13

E87495

prolyl oligopeptidase family protein [imported] - Caulobacter crescentus

C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C/Accession: E87495

R/Nerman, W.C.; Feilblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K

n, U.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of Caulobacter crescentus.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: E87495

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-683 <STO>

A/Cross-references: GB:AE005673; NID:913423453; PIDN:AAK3961.1; GSPDB:GN00148

A/Genetics:

A/Genetics: CC1986

Query Match 11.7%; Score 226.5; DB 2; Length 683;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 ; Search time 5.71145 Seconds

(without alignments)
2964.152 Million cell updates/sec

Title: US-10-070-464-7

Sequence: 1 EEDARSAGVATFVLQGEFDR.....RAVAGAPVTLMIFDYDGYT 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	16.6	765	1 DPP4_FELCA	Q9n217 felis silve
2	319	16.5	765	1 DPP4_BOVIN	P81425 bos taurus
3	315	16.3	766	1 DPP4_HUMAN	P27487 homo sapien
4	312.5	16.2	793	1 YEAB_SCHPO	Q14073 schizosacch
5	309	16.0	760	1 DPP4_MOUSE	P28843 mus musculu
6	297.5	15.4	767	1 DPP4_RAT	P14740 rattus norv
7	280.5	14.5	818	1 DAP2_YEAST	P18962 saccharomyc
8	259	13.4	761	1 SEPR_MOUSE	P97371 mus musculu
9	257	13.3	760	1 SEPR_HUMAN	Q12884 homo sapien
10	250	12.9	931	1 ST13_YEAST	P33894 saccharomyc
11	221	11.4	863	1 DPP6_BOVIN	P42659 bos taurus
12	212	11.0	859	1 DPP6_RAT	P46101 rattus norv
13	211	10.9	804	1 DPP6_MOUSE	Q92218 mus musculu
14	203	10.5	865	1 YDZE_HUMAN	P42658 homo sapien
15	197.5	10.2	853	1 YDZE_SCHPO	Q9p769 schizosacch
16	196	10.1	657	1 YUXI_BACSU	P39839 bacillus su
17	134.5	7.0	732	1 ACPH_HUMAN	P13798 homo sapien
18	131.5	6.8	732	1 ACPH_RAT	P13676 rattus norv
19	126.5	6.5	689	1 PPCE_AERHY	Q06903 aeromonas h
20	126.5	6.5	732	1 ACPH_PIG	P19205 sus scrofa
21	118	6.1	436	1 PAPA_CAVPO	P70683 c platelet
22	112.5	5.8	835	1 NLG3_HUMAN	Q8n124 homo sapien
23	110.5	5.7	836	1 NLG3_RAT	Q62888 rattus norv
24	106	5.5	705	1 PPCE_FLAAME	P27195 flavobacter
25	106	5.5	710	1 PPCE_PIG	P23687 sus scrofa
26	105	5.4	722	1 Y174_RICPR	P81171 rickettsia
27	103.5	5.4	705	1 PPCE_FLAAME	P27067 flavobacter
28	102.5	5.3	489	1 PNBA_BACSU	P37867 bacillus su
29	100	5.2	710	1 PPCE_BOVIN	Q9xta2 bos taurus
30	98.5	5.1	710	1 PPCE_MOUSE	Q9xtu6 mus musculu
31	98.5	5.1	444	1 PAPA_CANPA	Q28262 c platelet
32	98	5.1	705	1 Y450_RAISN	P55656 rhizobium s
33	96	5.0	710	1 PPCE_HUMAN	P48147 homo sapien

34	95	4.9	740	1 YL31_CABEL	P34422 caenorhabdi
35	95	4.9	848	1 NLG3_RAT	Q62889 rattus norv
36	94.5	4.9	840	1 NLG3_HUMAN	Q9n217 homo sapien
37	94.5	4.9	3354	1 CADN_HUMAN	Q9h251 homo sapien
38	94	4.9	440	1 PAPA_MOUSE	Q60963 m platelet
39	94	4.9	848	1 NLG3_HUMAN	Q9n217 homo sapien
40	93.5	4.8	419	1 CARV_CANAL	P10977 candida alb
41	93	4.8	367	1 DUS1_MOUSE	P28563 mus musculu
42	92.5	4.8	563	1 LIP2_GEOCN	P22394 geotrichum
43	92.5	4.8	843	1 NLG3_RAT	Q62765 rattus norv
44	92	4.8	816	1 NLG3_HUMAN	Q8n064 homo sapien
45	92	4.8	825	1 NLG3_MOUSE	Q8b5ms mus musculu

ALIGNMENTS

```

RESULT 1
DPP4_FELCA
ID DPP4_FELCA STANDARD; PRT; 765 AA.
AC Q9n217;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26).
GN DPP4 OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RX MEDLINE=20094000; PubMed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA Mikami T., Takahashi E.;
RT Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT activation antigen CD26 homologue.";
RL Immunogenetics 50:366-368(1999).
CC
CC -I- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline (By similarity).
CC -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -I- SUBUNIT: Homodimer (By similarity).
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -I- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -I- SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AB023952; BAA92344.1; -.
CC MEROPS; S09.003; -.
CC InterPro; IPR002469; DPPIV N term.
CC InterPro; IPR001375; Peptidase_S9.
CC InterPro; IPR002471; Prol_endopep_ser.
CC InterPro; IPR000379; Ser_ester_site.
CC Pfam; PF00930; DPPIV N term; 1.
CC PROSITE; PS00708; PRO_ENDOPP SER; 1.
CC Hydroxylase; Aminopeptidase; Dipeptidase; Serine peptidase;
CC Transmembrane; Glycoprotein; Signal-anchor.
CC CHAIN
1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM

```

FT CHAIN 38 765 (MDPP).
 FT DOMAIN 1 6 (SDPP) (BY SIMILARITY).
 FT TRANSMEM 7 29 CYTOPLASMIC (POTENTIAL).
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 30 765 EXTRACELLULAR (POTENTIAL).
 FT ACT SITE 629 629 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 765 AA; 88213 MF; 3EFCE98A22B175D9 CRC64;

Query Match 16.6%; Score 320; DB 1; Length 765;
 Best Local Similarity 25.1%; Pred. No. 1.1e-19;
 Matches 125; Conservative 59; Mismatches 156; Indels 158; Gaps 22;

QY 1 EEDARAGATFLQRE-FDRYSGYWCCKAETTPSGGKILRLYENDESEVEITH--- 56
 DB 189 EENAIYNGIADWYEEIFSAVSALWMSPGTF-----LAVAQFDQVPLEIEYGF 239
 QY 57 VTSPLMLETRRADSFRPKGTANPKV-----TFKK 86
 DB 240 YDESEIQYMTMRIRPKAGANPTVKLVITKDNLNPNATSVETPPAAMLTGDYIL 299
 QY 87 SEIMIDAGRIIVDEVRLV-----YFEGTK---DSPLEHLYV-----VSYPNG 129
 DB 300 CVTANABERISLQWIRKIONYSVMDIRYNNSTGKMISSAQEHHEMSTGQVGRFRA 359
 QY 130 EYTRILTD-----RGYSHSCCI---SQHCF-----FISKYSN 158
 DB 360 EHFHTSDGNEFYKIIISNEDGYHICFQIDKDCFTITKAMEVIGIALTTDYIYSN 419
 QY 159 QKN--PHCVSLYKLS-----SPEDDPTCK-----TKEFATILDSAGP-L 195
 DB 420 EYKMGPGGRNLTKIQLNDYTKVACLSCELKPE---RCQYYSVSPGSEAKYVQLRCSGPEL 476
 QY 136 PNYT-----PEIFSFESTTGT-----LYGMLYKPHDLP 226
 DB 477 PLYTHRSSNDELERLVLEJNSALDKMLQEVQPSKDLFIILNETKFWYQMLPPH-FDT 535
 QY 227 GKQYPTVLEFYGGPOVAVNNRFKGVKVRNLN---TLASLGYVVV-IDNRGCHRGKF 282
 DB 536 SKKYELLIDVAGPQSQ-----KADALFRIMWATYLASTENITVASFGRSRGYGDIT 589
 QY 283 ECAFFYKMGQIIRIDQVEGLQYIAGRYPIDLDVAGIHMSYGYGLSIMALMQRSIDIFRV 342
 DB 590 MEAVNRRIQTEFEVEQIEAARQF-SKMGFVDDKRIALIMQMSYGVYVTSVGLGAGSGVRC 648
 QY 343 AIAGAPVTLMTFYDTGYT 360
 DB 649 GIAVAPVSRWEYDSVYT 666

RESULT 2

DEP4_BOVIN
 ID _DP4_BOVIN STANDARD; PRT; 765 AA.
 AC P81425; Q8MMG8; (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (Adenosine deaminase complexing protein) (ADCP-I)

DE (Activation molecule 3) (ACT3) (WCI0).
 GN DPP4 OR CD26.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovine; Bos.
 NCBI_TaxID=9913;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphocytes;
 RX MEDLINE=22067734; PubMed=12073152;
 RA Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,
 RT Bohach G.A.,
 RT "Molecular characterization of bovine CD26 upregulated by a
 RT staphylococcal superantigen."
 RL Immunogenetics 54:216-220 (2002).
 [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
 RC TISSUE=Thymus;
 RX MEDLINE=22021197; PubMed=11981836;
 RA Giddon D.R., Howard C.J.;
 RT "CD26 is expressed on a restricted subpopulation of dendritic cells in
 RT vivo."
 RL Eur. J. Immunol. 32:1472-1481 (2002).
 RN [3]
 RP SEQUENCE OF 1-24.
 RC TISSUE=T-cell;
 RX MEDLINE=21482004; PubMed=11598101;
 RA Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,
 RA Naessens J., Bohach G.A.;
 RT "Identity of activation molecule 3 on superantigen-stimulated bovine
 RT cells is CD26."
 RL Infect. Immun. 69:7190-7193 (2001).
 RN [4]
 RP SEQUENCE OF 537-546.
 RC TISSUE=Kidney;
 RX MEDLINE=98293306; PubMed=9629661;
 RA Ben-Shoshan I., Parola A.H.;
 RT The CP-1 subunit of adenosine deaminase complexing protein from calf
 RT kidney is identical to human, mouse, and rat dipeptidyl peptidase
 RT IV."
 RL Comp. Biochem. Physiol. 119B:289-292 (1998).
 CC -!- FUNCTION: Removes N-terminal dipeptides sequentially from
 CC polypeptides having unsubstituted N-termini provided that the
 CC penultimate residue is proline. Binds and regulates the activity
 CC of ADA.
 CC -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-
 CC xcc, from a polypeptide, preferentially when xbb is pro, provided
 CC xcc is neither pro nor hydroxyproline.
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
 CC a soluble form (By similarity).
 CC -!- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and
 CC several immune system tissues.
 CC -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
 CC by proteolytic processing (By similarity).
 CC -!- SIMILARITY: Belongs to peptidase family S9B.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF461806; AAL67836.1; -
 CC EMBL; AY056834; AAL23628.1; -
 CC MEROPS; S09.003; -
 CC InterPro; IPR002471; ProL_endopep_ser.
 CC InterPro; IPR000379; Ser_estrs_site.
 CC Pfam; PF00930; DPPIV_N_term; 1.
 CC Pfam; PF00326; Peptidase_S9; 1.

DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 KM Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;.
 KM Transmembrane; Glycoprotein; Signal-anchor.
 FT CHAIN 1 765
 FT CHAIN 1 765
 FT CHAIN 38 765
 FT CHAIN 1 765
 FT TRANSMEM 7 29
 FT DOMAIN 1 6
 FT DOMAIN 30 765
 FT ACT_SITE 629 629
 FT ACT_SITE 707 707
 FT ACT_SITE 739 739
 FT CARBOHYD 84 84
 FT CARBOHYD 91 91
 FT CARBOHYD 149 149
 FT CARBOHYD 218 218
 FT CARBOHYD 228 228
 FT CARBOHYD 271 271
 FT CARBOHYD 280 280
 FT CARBOHYD 320 320
 FT CARBOHYD 392 392
 FT CARBOHYD 495 495
 FT CARBOHYD 684 684
 SQ SEQUENCE 765 AA; 88369 MW; B2165421F43B116 CRC64;
 Query Match 16.5%; Score 319; DB 1; Length 765;
 Best Local Similarity 24.5%; Pred. No. 1.3e-19;
 Matches 125; Conservative 56; Mismatches 147; Indels 182; Gaps 22;
 QY 1 EEDARSGAVTFLVQEE-FDRYSGYVWCPEKETTTPSGKTLRLIYENDSEVEIIT-- 56
 DB 189 KKVIVNGITDWTVEEVSFAVSLMWSPWSTP-----LAVAFNDTEVPLEIYSF 239
 QY 57 VTSPLMETERRADSPRYKGTANPKVTFKKSEIM-----IDAEGRILVDVRYRLVYEGT 111
 DB 240 YSPDESLOYPKTVKIPYKAGAVNPTIFEPVNVISLSLSPINATISQQLVPPGSVLIG---- 295
 QY 112 KDSPLREHLVYVSVMNGEVT-----RLDGRYSHSCCISQHCDF-----FLSKYSNQ- 159
 DB 296 ----DHYLCDVTWTEERISLQWLRI--QNVS-----IMDIDCDYDSTGRWISSVGRQH 344
 QY 160 -----KNHCY-----SLYKLSPP-----DPTCK--TYEPVATT 188
 DB 345 IEISTTGWGRFRPAEHPFTSDNSFYKIIISNEGYKHICHFQTDKRNCTFTKGAWEVI 404
 QY 189 ----- 188
 DB 405 GIEALNSDYLYISNEYKMPGARNLYKIQLNDYTKYTCLSCHEINPRCOYVSYSQGA 464
 QY 189 ----LDSAGP-LPDYT-----PPEIFSFESTTGTTL- 214
 DB 465 KYVQLRCGSGPLDYTLHNSNDKELEVLNNSDDLQVLDVQVQPSKPLFIHLHGTFW 524
 QY 215 YGMLVYRPHDQPKKYPVTVFIYGGPOVQVNNRFRKVKFRNL--TLASLGVVVV-1 270
 DB 525 YQMLPPEH-FDKSKTKPPLLEVYAGPSQ-----FADALFRINMATVYLASTENITVASF 577
 QY 271 DNEGSGRGLKEFGAEFYKMGQLEIDQVEGLQYASRYFIDLDRIYGVMSYGGVLTSL 360
 DB 578 DNGSGYGGDKIMHAIINRRIKGTREVEQIEATRQF-SKMGFVDDKRIATIMGWSYGGVITS 636
 QY 331 MALMQRSDIFRVAIAGAPVTLMIFPDYTGTT 360
 DB 637 MVLGAGSGVFKCGIAVAVPSKMEYDSVVT 666
 RESULT 3
 ID DP4 HUMAN STANDARD; PRT; 766 AA.
 AC P27487;
 DT 01-AUG-1992 (Rel. 23, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (TP103) (Adenosine deaminase complexing protein-2) (ADABP)
 DE (ADABP)
 GN DPP4 OR ADCP2 OR CD26.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_Taxid=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92329551; PubMed=1352704;
 RA Msumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a serine proteinase on the cell surface.";
 RL Biochim. Biophys. Acta 1131:333-336(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95012454; PubMed=7927537;
 RA Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
 RT "Genomic organization, exact localization, and tissue expression of the human CD26 (dipeptidyl peptidase IV) gene.";
 RL Immunogenetics 40:331-336(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Peripheral blood;
 RX MEDLINE=92352530; PubMed=1352530;
 RA Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J., Dahlberg H.N., Schlossman S.F., Morimoto C.;
 RT "Cloning and functional expression of the T cell activation antigen CD26.";
 RL J. Immunol. 149:481-486(1992).
 RN [4]
 RP ERRATUM.
 RX MEDLINE=93171637; PubMed=8094732;
 RA Tanaka T.;
 RL J. Immunol. 150:2090-2090(1993).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Kausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Heien F., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F., Datchenko U., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunnarane P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Huliy S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalins D.E., Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [6]
 RP SEQUENCE OF 1-551 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=92165847; PubMed=1347043;
 RA Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C., Trolot P., Barbat A.;
 RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer cell lines HT-29 and Caco-2. Cloning of the complete human coding sequence and changes of dipeptidyl peptidase IV mRNA

RT levels during cell differentiation";
 RL J. Biol. Chem. 267:4824-4833(1992).
 RN [7]
 RP SEQUENCE OF 545-766 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=91024044; PubMed=1977364;
 RA Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;
 RT "Isolation of a cDNA probe for the human intestinal
 RT dipeptidylpeptidase IV and assignment of the gene locus DPP4 to
 RT chromosome 2.";
 RL Ann. Hum. Genet. 54:191-197(1990).
 RN [8]
 RP SEQUENCE OF 1-31 FROM N.A.
 RX MEDLINE=96067599; PubMed=7487939;
 RA Beehm S.K., Gum J.R. Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
 RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific
 RT regulation from a TATA-less GC-rich sequence characteristic of a
 RT housekeeping gene promoter.";
 RL Biochem. J. 311:835-843(1995).
 RN [9]
 RP PARTIAL SEQUENCE.
 RC TISSUE=Kidney;
 RX MEDLINE=93210468; PubMed=8096237;
 RA Morrison M.E., Vijayasatadhi S., Engelstein D., Albino A.P.,
 RA Houghton A.N.;
 RT "A marker for neoplastic progression of human melanocytes is a cell
 RT surface ectopeptidase.";
 RL J. Exp. Med. 177:1135-1143(1993).
 CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
 CC polypeptides having unsubstituted N-termini provided that the
 CC penultimate residue is proline. Plays a role in T cell activation.
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
 CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
 CC Xcc is neither Pro nor hydroxyproline.
 CC -1- SUBUNIT: Homodimer or heterodimer with Seprase (FAP).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
 CC a soluble form.
 CC -1- PPM: The soluble form (SDPP) derives from the membrane form (MDPP)
 CC by proteolytic processing.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -1- DATABASE: NAME=PROM; NOTE=CD guide CD26 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd26.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U13735; AAB60646.1; JOINED.
 DR EMBL: U13710; AAB60646.1; JOINED.
 DR EMBL: U13711; AAB60646.1; JOINED.
 DR EMBL: U13712; AAB60646.1; JOINED.
 DR EMBL: U13713; AAB60646.1; JOINED.
 DR EMBL: U13714; AAB60646.1; JOINED.
 DR EMBL: U13715; AAB60646.1; JOINED.
 DR EMBL: U13716; AAB60646.1; JOINED.
 DR EMBL: U13717; AAB60646.1; JOINED.
 DR EMBL: U13718; AAB60646.1; JOINED.
 DR EMBL: U13719; AAB60646.1; JOINED.
 DR EMBL: U13720; AAB60646.1; JOINED.
 DR EMBL: U13721; AAB60646.1; JOINED.
 DR EMBL: U13722; AAB60646.1; JOINED.
 DR EMBL: U13723; AAB60646.1; JOINED.
 DR EMBL: U13724; AAB60646.1; JOINED.
 DR EMBL: U13725; AAB60646.1; JOINED.
 DR EMBL: U13726; AAB60646.1; JOINED.
 DR EMBL: U13727; AAB60646.1; JOINED.
 DR EMBL: U13728; AAB60646.1; JOINED.
 DR EMBL: U13729; AAB60646.1; JOINED.
 DR EMBL: U13730; AAB60646.1; JOINED.

DR EMBL: U13731; AAB60646.1; JOINED.
 DR EMBL: U13732; AAB60646.1; JOINED.
 DR EMBL: U13733; AAB60646.1; JOINED.
 DR EMBL: U13734; AAB60646.1; JOINED.
 DR EMBL: M74777; AAA51943.1; -.
 DR EMBL: BC013329; AAH13329.1; -.
 DR EMBL: M80536; AAA52308.1; -.
 DR EMBL: X60708; CAA43118.1; -.
 DR EMBL: S79876; AAB35614.1; -.
 DR PIR: S24313; CDH26.
 DR PDB: 1N1M, 07-JAN-03.
 DR MEROPS: S09.003; -.
 DR Genew: HGNC:3009; DPP4.
 DR MIM: 102720; -.
 DR GO: GO:0004274; F:dipeptidyl-peptidase IV activity; TAS.
 DR GO: GO:0005209; F:Plasma protein; TAS.
 DR GO: GO:0006955; P:Immune response; TAS.
 DR InterPro: IPR002469; DPP4V N term.
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR002471; Prol_endopep_ser.
 DR InterPro: IPR000379; Ser_estrs_site.
 DR Pfam: PF00930; DPP4V N_term; 1.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
 KW Transmembrane; Glycoprotein; Signal anchor; 3D-structure
 FT CHAIN 1 766
 FT DOMAIN 39 766
 FT TRANSSEM 7 28
 FT DOMAIN 6 766
 FT ACT SITE 630 630
 FT ACT SITE 708 708
 FT ACT SITE 740 740
 FT CARBOHYD 85 85
 FT CARBOHYD 92 92
 FT CARBOHYD 150 150
 FT CARBOHYD 219 219
 FT CARBOHYD 229 229
 FT CARBOHYD 281 281
 FT CARBOHYD 321 321
 FT CARBOHYD 520 520
 FT CARBOHYD 685 685
 FT CONFLICT 6 6
 FT CONFLICT 7 7
 FT CONFLICT 437 437
 FT CONFLICT 557 557
 FT CONFLICT 663 663
 Query Match 16.3%; Score 315; DB 1; Length 766;
 Best local Similarity 23.4%; Pred. No. 2,9e-19;
 Matches 120; Conservative 59; Mismatches 146; Indels 188; Gaps 22;
 QY 1 EEDARSAGVATFVLQEE-FPRYSGYWCPCPAETTPSGKRLRLIYEANDSEVEIHH---56
 DB 190 KEDIIYNGIDTWYBEEVFSAYSALWWSFNGTF-----LATAQFNTFVFLLEYSF 240
 QY 57 VTSPLMETRRADSFYPKTGTANPKYTF-----KMSIMIDAGRIIVDEVRR 104
 DB 241 YSDESIOLPKTIVAPFKAGAVPTVFFVYVNTDSLSVYVNAISIQTAASMLIG-----296
 QY 105 LVYFEGTKSPLEHLHYVVSYNPGEVT-----RLDRGYSHSC-----CI--SQ 147
 DB 297 -----DHYLCDTVMTAQRISIQWLRRIQYVYMDICDYDESSGRNMCIVARQ 344
 QY 148 HCDF-----FISKXSNQKNPCV-----SLVLTSSPED-----DPTCKTGFWA 186
 DB 345 HEMSTTGWVGRF-RSEPHFTLDGNSFYKILSNBGRHRCYFOIDKKOCTETTKTWE 403
 QY 187 TI-----188

Db 404 VIGIEALTSYLYIYSNEYGMPCGRNLKYLQSDTYTKVTCLSCEINPERCQYYSFSK 463
 QY 189 -----LDSAGP-LPDYT-----PPIFSF--ESTTG 211
 Db 464 EAKYQQLRCGSGPLPLTLTHSSVNDKGLFVLENSALDKMLQVQMSKLDPIINETK 523
 QY 212 FTLYGMLYKHHDLPQKKYFTVLFYGGPOVQVNNRFKGVKFRLN---TLASLGYVV 268
 Db 524 F-WYQWILPQH-FDKSKKYPPLLDVYAGPCSQ-----KADTVFRINWATYLASTENIIV 575
 QY 269 V-IDNQSCHRGKLFEGAFYKMGQIEIDPOVSLQYLAIRYPIIDLRGIGHWSYGGY 327
 Db 576 ASFDGKSGYQGDIMHAINRRLGTFEVEDQIEARQF-SKMGFVMDKRIAIGWSYGGY 634
 QY 328 LSLMALMQRSDIFRVALAGAPVTLMIPTDGYT 360
 Db 635 VTSMWLGSGSGVFKCGIAVAPVRWEYDVSYYT 667

RESULT 4
 YEAR SCHPO STANDARD; PRT; 793 AA.
 ID YEAR SCHPO
 AC 014073;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative dipeptidyl aminopeptidase C2B11.08 (EC 3.4.14.-).
 GN SPACUNK4.08 OR SPAC2B11.08.
 OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NX NCB1_TaxId=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Genies S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Wolkaert G., Aert R., Robben J., Grymouprez B.,
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer C., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreno S., Gloux S., Lelaurie V., Mottier S.,
 RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shporkovskii G.V., Ussery D., Barrett B.G., Nurse P.,
 RA "the genome sequence of Schizosaccharomyces pombe";
 RT Nature 415:871-880(2002).
 RL
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
 CC vacuoles (By similarity).
 CC -I- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 Db EMBL: AL031180; CAA20138.1; -.
 DR PIR: T41703; T41703.
 DR GeneDB_Spombe; SPACUNK4.08; -.
 DR InterPro: IPR002469; DPIP_V_N_term.
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR002471; ProL_endopep_ser.
 DR InterPro: IPR000379; Ser_estrg_site.
 DR Pfam: PF00930; DPIP_V_N_term; 1.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR Hypothesis: P800708; PRO_ENDOPEP_SER; 1.
 KM Hypothesis: P800708; PRO_ENDOPEP_SER; 1.
 KM Serine protease: Transmembrane; Glycoprotein; Signal-anchor.
 FT DOMAIN 1 24
 FT TRANSMEM 25 45
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT LUMENAL (POTENTIAL).
 FT ACT SITE 647 647
 FT ACT SITE 722 722
 FT ACT SITE 755 755
 FT CARBOHYD 101 101
 FT CARBOHYD 136 136
 FT CARBOHYD 166 166
 FT CARBOHYD 246 246
 FT CARBOHYD 299 299
 FT CARBOHYD 303 303
 FT CARBOHYD 324 324
 FT CARBOHYD 346 346
 FT CARBOHYD 377 377
 FT CARBOHYD 407 407
 FT CARBOHYD 535 535
 FT CARBOHYD 761 761
 SQ SEQUENCE 793 AA; 91304 MW; 20870797F231463 CRC64;
 Query Match 16.2%; Score 312.5; DB 1; Length 793;
 Best Local Similarity 22.9%; Pred. No. 5e-19;
 Matches 114; Conservative 62; Mismatches 149; Indels 173; Gaps 17;
 QY 8 GVATFVLQEE-PDRISGVYWCRAETTPGGKILRLYEENDESEVEIIV----- 57
 Db 214 GLTDWYEEVYSSPETIWWSPDS-----KIAFLKINSEIPTYHYPIYTAELDP 264
 QY 58 TSPMLETRADSRPYKCTGTANPKYTFKSEI----- 89
 Db 265 SLPEPDYKQMAIKIPKPNPNPVSILFVADLNSNASSNFSIMHNEPLAPVQVNLWNV 324
 QY 90 -----MIDAEGRIL----- 98
 Db 325 TSSVLVQFTNRNSTCITATLDTLTKSITVTTECLBEGWYVQSAKAPFLNLSLWEN 384
 QY 99 -----VDEVRRL-----VYEGTKDSP 115
 Db 385 WSDGYFDLADDDYHNLAFIPNGSSPIYLSGADVDTGPIHIDGFGNVYLAFLTKDS 444
 QY 116 LEEHLVYVSVNGEYTRLTDRGYS-----HSCISQCHDPEFIKSYNQKPHCVSLYKLS 172
 Db 445 TERHLIYYS-LDTLEIYGITDGEDEGYSTFSFGPDYVINYHGPDPV-----WQELKS 499
 QY 173 PEDDPTCKTEPMATILDSAGPLPDYTPPEI-----FSFESTTGFTLYGMLYKPHDQPEK 228
 Db 500 TKDQXCYSLFETNSRLKQ-----LSSITLPSVAYGLTFNDTL-FNF--MERPRPNVDVK 553
 QY 229 KYPYTLFYGGGQVQVUNNRFQVGYFRNLATLSIG-----YVVVVIDNRSGCHRGKPF 282
 Db 554 KIPVYFFAYVGGGQV-----ATLFFVVDPAVYASHPDPEFFIVTLIDRGRTGFNGNAF 607
 QY 283 EGAFYKMGQIEIDPOVGLQYLAIRYPIIDLRGIGHWSYGGYLSLMAIMQSRDIFRY 342
 Db 608 RYSVSRHGHGMESTDQAGAGFMAD-LPFVDENHVGWMSYGYLTILKTL-ETQDVFSY 665
 QY 343 AIAGAPVTLMIPTDGYT 360

Db 666 GMAVAPYTDWRLYDSVYT 683

RESULT 5
DPP4_MOUSE STANDARD; PRT; 760 AA.
AC P28843;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (Thymocyte-activating molecule) (THAM).
GN DPP4 OR CD26.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Thymus;
RX MEDLINE=92129288; PubMed=1370813;
RA Marguet D.A., Bernard A.-M., Vivier I., Darmon D., Naquet P.,
RT Pierres M.;
RT "cDNA cloning for mouse thymocyte-activating molecule. A
RT multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
RT subgroup of serine proteases.";
RT J. Biol. Chem. 267:2200-2206(1992).
RN [2]
RP REVISIONS.
RA Marguet D.A.;
RN Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.A; TISSUE=Liver;
RX MEDLINE=95092780; PubMed=7999781;
RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
RT "Structure of the mouse dipeptidyl peptidase IV (CD26) gene.";
RT Biochemistry 33:15204-15214(1994).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Collins B., Buetow K.H., Scheen C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stamenkovic M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshlyuk S., Abramson R.D., Mallahy S.J.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallahy S.J.,
RA Bosak S.A., McMan P.J., McKernan K.J., Malek A.M., Gay L.J., Hulyk S.W.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Gittwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherich A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP SEQUENCE OF 1-20.
RX MEDLINE=91302787; PubMed=1712807;
RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,
RA Bernard A.-M., Gorvel J.-P., Pierres M.;
RT "Evidence that thymocyte-activating molecule is mouse CD26
RT (dipeptidyl peptidase IV).";
RT J. Immunol. 147:447-454(1991).
CC -I- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline.

CC -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -I- SUBUNIT: Homodimer.
CC -I- SUPRACELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form.
CC -I- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing.
CC -I- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and thus statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58384; CAA41274.1; -
DR EMBL: U12620; AAA82213.1; -
DR EMBL: U12599; AAA82213.1; JOINED.
DR EMBL: U12600; AAA82213.1; JOINED.
DR EMBL: U12601; AAA82213.1; JOINED.
DR EMBL: U12602; AAA82213.1; JOINED.
DR EMBL: U12603; AAA82213.1; JOINED.
DR EMBL: U12604; AAA82213.1; JOINED.
DR EMBL: U12605; AAA82213.1; JOINED.
DR EMBL: U12606; AAA82213.1; JOINED.
DR EMBL: U12607; AAA82213.1; JOINED.
DR EMBL: U12608; AAA82213.1; JOINED.
DR EMBL: U12609; AAA82213.1; JOINED.
DR EMBL: U12610; AAA82213.1; JOINED.
DR EMBL: U12611; AAA82213.1; JOINED.
DR EMBL: U12612; AAA82213.1; JOINED.
DR EMBL: U12613; AAA82213.1; JOINED.
DR EMBL: U12614; AAA82213.1; JOINED.
DR EMBL: U12615; AAA82213.1; JOINED.
DR EMBL: U12616; AAA82213.1; JOINED.
DR EMBL: U12617; AAA82213.1; JOINED.
DR EMBL: U12618; AAA82213.1; JOINED.
DR EMBL: U12619; AAA82213.1; JOINED.
DR EMBL: BC022183; AAH22183.1; -
DR MEROPS: S09.003; -
DR MGD: M61.949194; DPP4.
DR InterPro: IPR002469; DPP4 N-term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR002471; PIOL_endopep_ser.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00930; DPP4V N-term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 760
FT FT
FT FT
FT CHAIN 37 760
FT FT
FT FT
FT DOMAIN 1 6
FT FT
FT FT
FT TRANSMEM 7 28
FT FT
FT FT
FT DOMAIN 29 760
FT FT
FT FT
FT ACT_SITE 624 624
FT FT
FT ACT_SITE 702 702
FT FT
FT ACT_SITE 734 734
FT FT
FT CARBOHYD 83 83
FT FT
FT CARBOHYD 90 90
FT FT
FT CARBOHYD 113 113
FT FT
FT CARBOHYD 213 213
FT FT
FT CARBOHYD 223 223
FT FT
FT CARBOHYD 315 315
FT FT
FT CARBOHYD 328 328
FT FT
FT CARBOHYD 514 514

FT	CONFLICT	332	332	T -> N (IN REF. 2).	
FT	CONFLICT	352	352	C -> V (IN REF. 2). <td></td>	
FT	CONFLICT	394	394	V -> D (IN REF. 2). <td></td>	
FT	CONFLICT	562	562	L -> F (IN REF. 2). <td></td>	
FT	CONFLICT	624	624	R -> Q (IN REF. 2). <td></td>	
SO	SEQUENCE	767 AA;	88003 MW;	AA17COCECF054652 CRC64;	
Query Match 15.4%; Score 297.5; DB 1; Length 767;					
Best Local Similarity 24.5%; Pred. No. 9.4e-18;					
Matches 121; Conservative 60; Mismatches 152; Indels 161; Gaps 22					
QY	8	GVATFVLQGE-FDRSGVWVCCKATTTSGGKILRIIYENDESEVEIHHVTSPLMETRR	66		
DB	195	GINDWVYEETIEGASALMW-----SPNGFIAYAGFNDGVPLIEFSYSSESQYQR	248		
QY	67	ADSFRRPKGTANPKVTF-----			
DB	249	TWMIPIPKAGVNPFTKFIYNTDSLSTTTTITMQITADASVTTGPHYLCVAMVSEDR	308		
QY	97	IIVDERRLVTF-----EGTDS-----PLEHLLVYVSIV	126		
DB	309	ISIQMLRKIQNSVMAICDYDKTILWMNCPTTQHIETISATGCGRRPAPRH-----FT	363		
QY	127	NPEGE-VTFLTLDR-GYSHSCCI-----SQHCFD-----FISKYSNQ-	159		
DB	364	SDGSSPYKIVSDKDGKHLICQGRKREQVCTFTGAMBEVISEALTSIDVLIYSNEY	423		
QY	160	KMPHCVSIVK-----LSPEDDPICK-----IKETMATILDSAGI-LPDTY-	199		
DB	424	KEMPGGRNLKIQQLDHTNKKCLSCDLPERCQYYSVLSKKAQYVOLGCGRGGLPYTL	483		
QY	200	-----PPEIFSFESTGTFT-----LYGMLYXPHDLOPEKKTP	231		
DB	484	HRSTDQKEDRLVEDNSALDKMLQDVMPKSKLDFIETNETRFWQXHLIPRH-FDKSKKTP	542		
QY	232	TVLFTIGGGQVQVLYNNRFGVYGFELN--TLASLGVVVV-IDNGSCIRGLKPEGARK	287		
DB	543	LRLDVYAGPSSQ-----KADAFELNATYTLASTENIIVASFGRSRGQGDKIMHAIN	596		
QY	288	YKMGQRLIDDQVEGL-QYLASRYDFLIDRVLGHGWSYGYTLSTMLMORSIDIFRVAIAG	346		
DB	597	KRLGTLEVEDQLEAARQFL--KMGFVDSKRFVAILMGWSYGYVTSMVLGSGVFKGIGIAY	654		
QY	347	APYTLMIIFYDTGTY	360		
DB	655	APVSRMEYIDSVYT	668		
RESULT 7					
DAP2_YEAST					
ID	DAP2_YEAST	STANDARD;	PRT;	818 AA.	
AC	P18962;				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DPAp B) (YSCV).				
GN	DAP2 OR YHR028C.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxId=4932;				
RN	[1]				
RX	SEQUENCE FROM N.A.				
RX	MEDLINE=89174971; PubMed=2647766;				
RA	Roberts C.J., Pohlig G., Rothman J.R., Stevens T.H.;				
RT	"Structure, biosynthesis, and localization of dipeptidyl				
RT	aminopeptidase B, an integral membrane glycoprotein of the yeast				
RT	vacuole.";				
RL	J. Cell Biol. 108:1363-1373(1989).				
RN	[2]				
RX	SEQUENCE FROM N.A.				
RC	STRAIN=S288c / AB972;				
RX	MEDLINE=94378003; PubMed=8091229;				

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten Y.,
 RA Kuchta T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
 RA Nham M., Rikkin L., Riles L., St Peter H., Trevaastis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
 RA Vainid M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 VII".
 RT Science 265:2077-2082(1994).
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
 CC VACUOLLES.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
 CC the European Bioinformatics Institute. There are no restrictions on ways
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X15484; CAA3512.1; -;
 CC EMBL; U10399; AAB6879.1; -;
 CC PIR; S46780; A30107.
 CC MEROPS; S09.006; -;
 CC SCD; S0001070; DAP2.
 CC GO; GO:0000329; C:vacuolar membrane (sensu Fungi); IDA.
 CC InterPro; IPR002469; DDP1V_N_term.
 CC InterPro; IPR001375; Peptidase_S9.
 CC InterPro; IPR002471; Prol_endopep_ser.
 CC InterPro; IPR000379; Ser_ester_site.
 CC Pfam; PF00930; DDP1V_N_term; 1.
 CC Pfam; PF00326; Peptidase_S9; 1.
 CC PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 CC Hydroxylase, Aminopeptidase; Dipeptidase, Serine protease;
 CC Transmembrane; Glycoprotein; Signal_anchor.
 CC CYTOPLASMIC (POTENTIAL).
 CC SIGNAL_ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC DOMAIN
 CC 1 29
 CC TRANSSEM 30 45
 FT FT DOMAIN 46 818
 FT ACT SITE 679 679
 FT ACT SITE 756 756
 FT ACT_SITE 789 789
 FT CARBOHYD 63 63
 FT CARBOHYD 79 79
 FT CARBOHYD 110 110
 FT CARBOHYD 139 139
 FT CARBOHYD 372 372
 FT CARBOHYD 392 392
 FT CARBOHYD 421 421
 FT CARBOHYD 738 738
 FT CONFLICT 83 83
 FT CONFLICT 125 125
 FT CONFLICT 182 188
 FT CONFLICT 200 200
 FT CONFLICT 366 375
 FT CONFLICT 808 818
 SQ SEQUENCE 818 AA; 93404 MW; 318F450445375BD3 CRC64;
 Query Match 14.5%; Score 280.5; DB 1; Length 818;
 Best Local Similarity 22.0%; Pred. No. 3e-16;
 Matches 114; Conservative 50; Mismatches 155; Indels 199; Gaps 19;
 QY 8 GVATFVLOER-FDHYSGYWMCPKALPTPSGGKILRIIYEENDESEV-EIHVTSPLMETR 65
 DB 234 GKPMVYEEVFEEDDKAAMSPFD-----YLAFKIDSEVGEFI--IPYVDD 281
 QY 66 RAD-----SRVYKGTGAP----- 80
 DB 282 EKDIYEWKRSIKYKSGTPNPHAEIWLWYSMKDGTSPHPRISGNKKDGLITETWYWGNG 341

QY 60 PMLETRRADSPRYKGTGNPKY-----TFK 85
 Db 245 P-----RTINIPYKAGAKNEVAVFIVDTTYPHHVGMVEVPEMTASSDYPFSTWLTWV 299
 QY 86 MSE----- 88
 Db 300 SSEVVCQWLKRVQNVSVLSICDPREDHNAMECPKQNHVEESTTGAGCFVSTPAFSQ 359
 QY 89 -----IMDAEG-----RIIVDEVRLV-----YFEGTKDS-----P 115
 Db 360 DATSYVKTFSDKDGKHIHYIKDVTENAIQITSGKWEALYIFRVQSLFYSSNEFGYF 419
 QY 116 LEHLVYVSYNPEGEVTRLDRGSHSCISQH-----CPEFISKYNSQKPHCVSLY-- 168
 Db 420 GRRIIVYISIGNSPSK-----CCTCHLRKRCQYITAFSFKAKYVALVCYGP 469
 QY 169 --KLSPEDDPYCKTEKFEWATILDSAGPLDPYTPPEI--FSESTYGTLYGMLYKPHDLQ 225
 Db 470 GLPISTLHDGRTDGEIQVLEENKELNSLNIOCPKYEIKKLKDGGLTFYKMLLPQFD 529
 QY 226 PGKKYPTVFLTYGGPQVLYNNRKYKRYRLNTLASL---GYVVVINDRSGCHRGK 281
 Db 530 RSKKYPPLLIQYGGPCQSVKS-----VPAVNMITYLAKSEGIIVIALVDGRGTAFGDK 583
 QY 282 FEGAFKYMGCIEIDQVEGLOYLASRYDFIDRVGIGHMSYGYSLMALMQRSDIER 341
 Db 584 FLNAVYKLGVEVVEDQLTAVRKTI-EMGFIDEBRIAWMSYGVYSIALASGLTLPFK 642
 QY 342 VAIGAPVTLMIFYDTGYT 360
 Db 643 CGIAVAVSSMEYASIYS 661

RESULT 9
 SEPR_HUMAN STANDARD; PRT; 760 AA.
 AC Q12884; Q00199; Q09998; Q0UID4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (integral membrane serine protease) (170-kDa melanoma membrane-bound gelatinase).
 DE FAP.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Fibroblast;
 RX MEDLINE=94261645; PubMed=7911242;
 RA Scanlan M.J., Raj B.K.M., Celivo B., Garin-Chesa P., Sanz-Moncasi M.P., Healey J.H., Old L.J., Rettig W.J.;
 RT "Molecular cloning of fibroblast activation protein alpha, a member of the serine protease family selectively expressed in stromal fibroblasts of epithelial cancers.";
 RT Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Melanoma;
 RX MEDLINE=97388251; PubMed=9247085;
 RA Goldstein L.A., Ghersi G., Pineiro-Sanchez M.L., Salamone M., Yeh Y., Plesate D., Chen W.-T.;
 RT "Molecular cloning of seprase: a serine integral membrane protease from human melanoma.";
 RT Biochim. Biophys. Acta 1361:11-19(1997).
 RN (3)
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND 511-518.
 RC TISSUE=Melanoma;
 RX MEDLINE=97218181; PubMed=9065413;

RA Pineiro-Sanchez M.L., Goldstein L.A., Dodt J., Howard L., Yeh Y., Chen W.-T.;
 RT "Identification of the 170-kDa melanoma membrane-bound gelatinase (seprase) as a serine integral membrane protease.";
 RT J. Biol. Chem. 272:7595-7601(1997).
 RN (4)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Melanoma;
 RX MEDLINE=20112818; PubMed=10644713;
 RA Goldstein L.A., Chen W.-T.;
 RT "Identification of an alternatively spliced seprase mRNA that encodes a novel intracellular isoform.";
 RT J. Biol. Chem. 275:2554-2559(2000).
 RN (5)
 RP SEQUENCE OF 192-208; 220-240 AND 510-521.
 RX MEDLINE=9437249; PubMed=7519584;
 RA Rettig W.J., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M., Garin-Chesa P., Healey J.H., Old L.J.;
 RT "Fibroblast activation protein: purification, epitope mapping and induction by growth factors.";
 RT Int. J. Cancer 58:385-392(1994).
 CC -1- FUNCTION: May have a role in tissue remodeling during development and wound healing, and may contribute to invasiveness in malignant cancers.
 CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and type IV collagen, but not native type I or type IV collagen. Does not cleave laminin, fibronectin, fibrin or casein.
 CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is inactive.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell surface lamellipodia, invadopodia and on shed vesicles.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=L;
 CC IsoId=Q12884-1; Sequence=Displayed;
 CC Note=Major isoform;
 CC Name=2; Synonyms=S, Truncated;
 CC IsoId=Q12884-2; Sequence=VSP_005367;
 CC -1- TISSUE SPECIFICITY: Fibroblast-specific.
 CC -1- INDUCTION: In fibroblasts at times and sites of tissue remodeling during development, tissue repair, and carcinogenesis.
 CC -1- PTM: N-glycosylated.
 CC -1- PTM: The N-terminus may be blocked.
 CC -1- SIMILARITY: Belongs to peptidase family 99B.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; U09278; AAB49652.1; -
 CC EMBL; U76833; AAC51668.1; -
 CC EMBL; AF007822; AAF21600.1; -
 CC MEROPS; S09.007; -
 CC GeneW; HGNC:3590; FAP.
 CC MIM; 600403; -
 CC InterPro; IPR002469; DPPIV N term.
 CC InterPro; IPR001375; Peptidase_S9.
 CC InterPro; IPR002471; Pro_endopep_ser.
 CC InterPro; IPR000379; Ser_estrs_site.
 CC Pfam; PF00930; DPPIV_N term; 1.
 CC Pfam; PF00326; Peptidase_S9; 1.
 CC PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 CC HydroLase; Protease; Serine protease; Transmembrane; Signal-anchor; Glycoprotein; Alternative splicing;
 CC DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 CC DOMAIN 26 760 EXTRACELLULAR (POTENTIAL).
 CC FT 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 624 624

FT ACT_SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPEC 1 521 Missing (in isoform 2).
 FT CONFLICT 207 207 /FTID=VSP_005367.
 FT CONFLICT 229 229 P -> A (IN REF. 2).
 FT CONFLICT 354 354 K -> T (IN REF. 2).
 FT CONFLICT 354 354 R -> T (IN REF. 2).
 SQ SEQUENCE 760 AA; 87820 MM; A0D34B4801B07BA CRC64;

Query Match 13.3%; Score 257; DB 1; Length 760;
 Best Local Similarity 22.7%; Pred. No. 2,9e-14;
 Matches 110; Conservative 68; Mismatches 158; Indels 148; Gaps 20;

QY 8 GVATFVLQSEF--DRYSGYWCPCRAETTPSGKTLRLTYEENDESEVEIHTVSPMLT- 64
 DB 195 GIDWYEEEMLEPTKA-LWMS-NGKFL-AAAEFNDKDIPIVLAISYGDQY 244
 QY 65 RRADSEPRYKVTGTANPKVTFKM-----SEIMIDA-----EGRI 97
 DB 245 PRINIPYPRAGAKNPVRFIFIDTTPAVVGQEVVPRAMIASDYPSMLTWIDERY 304
 QY 98 IYDEVKRLVYFE-----GTQDSP-LEHLY-----VVSYPGE 130
 DB 305 CLQMLKRVQNVSVLSICDFREDWQTDCKPTQHEIESRTGAGCFVSPSPSYDAISY 364
 QY 131 VTRLTDR-GYSHSCCI-----SQHDFEIS-----KYSQOKRP 162
 DB 365 YKIFSDKGYKIHVYKQIVENAIOITSGKWEINFRVYQDSLFYSSNEFEYPERNNI 424
 QY 163 HCVSIVLYKSSPEDDPTCKTE---FMATILDSA-----GP-LPDT- 199
 DB 425 YRISIGSPYPSKCVTCHLRKRCQYRTASFSQYAKYVALVCGPEIPISLTDGRTDGE 484
 QY 200 -----PPRIFSESTTGFTLYMLYKPHDLQGGKYPTVLFYGGP 240
 DB 485 IKILEENKELENALKNIOLPKEIKKLEVEDEITLWYKMLIPQDFRSKKYPLLIQYGGP 544
 QY 241 QVQVNNRPFKGVYFELNTLASI---GYVVVVINDRGSCCHGLKREGAFKRMQIETD 296
 DB 545 CSQASRS-----VRAVWISYTLASKEGVNIALVDRGAFGSDKLAVYKRLGYEVE 598
 QY 297 DQVEGLQYLASRYDFIDRVGIGHWSYGYLSLMMQSRDIFRVALAGAPVTLIMFYD 356
 DB 599 DQITAVRKPI-EMGFIDEKRIAIWMSYGYVSSIALAGTGLFKGGINAVAVSSMEYTA 657
 QY 357 TGYT 360
 DB 658 SVYT 661

RESULT 10
 STL3_YEAST STANDARD; PRT; 931 AA.
 AC P33894;
 DT 01-FEB-1994 (Rel. 28; Created)
 DT 01-FEB-1994 (Rel. 28; Last sequence update)
 DT 28-FEB-2003 (Rel. 41; Last annotation update)
 DE Dipeptidyl aminopeptidase A (EC 3.4.14.-) (DPAP A) (YSCIV).
 GN STL3 OR YCII OR YOR219C OR YOR50-9.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95066382; PubMed=7975897;
 RA Anna-Miriola S.S., Herskowitz I.,
 RT "Isolation and DNA sequence of the STL3 gene encoding dipeptidyl

RT aminopeptidase.";
 RL yeast 10:801-810(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AB320;
 RA Flanagan C.A., Thorne J.;
 RT "STL3";
 RL (in) Getting M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.);
 RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
 Press, Oxford (1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=96437977; PubMed=8840505;
 RA Galsison F., Dujon B.;
 RT "Sequence and analysis of a 33 kb fragment from the right arm of
 chromosome XV of the yeast Saccharomyces cerevisiae.";
 RL yeast 12:877-885(1996).
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
 ALPHA-FACTOR PRECURSOR.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
 VACUOLES.
 CC -1- SIMILARITY: Belongs to peptidase family 99B.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; L21944; AAA35119.1; -
 CC EMBL; U08230; AAA17897.1; -
 CC EMBL; X92441; CAA3182.1; -
 CC EMBL; Y75127; CAA9437.1; -
 CC FIR; A49737; A49737.
 CC MEROPS; S09.005; -
 CC DR SGD; S0005745; STR13.
 CC DR GO; GO:0005802; C:Golgi trans-face; IDA.
 CC DR GO; GO:004177; P:aminopeptidase activity; IDA.
 CC DR GO; GO:007323; P:peptide pheromone maturation; IDA.
 CC DR InterPro; IPR002469; DPTIV_N term.
 CC DR InterPro; IPR001375; Peptidase_S9.
 CC DR InterPro; IPR002471; Prol endopep. ser.
 CC DR InterPro; IPR000379; Ser_cstrs_site.
 CC DR Pfam; PF00330; DPTIV_N term; 1.
 CC DR Pfam; PF00326; Peptidase_S9; 1.
 CC DR PROSITE; PS00708; PRO_ENDOPEP_SER; FALSE NEG.
 CC KW Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
 CC KW Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
 CC FT DOMAIN 1 119
 CC FT TRANSMEM 120 140
 CC FT DOMAIN 141 931
 CC FT ACT_SITE 785 785
 CC FT ACT_SITE 863 863
 CC FT ACT_SITE 896 896
 CC FT CARBOHYD 377 377
 CC FT CARBOHYD 814 814
 CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 931 AA; 107200 MM; 81AF70094093C023 CRC64;

Query Match 12.9%; Score 250; DB 1; Length 931;
 Best Local Similarity 24.3%; Pred. No. 1.5e-13;
 Matches 101; Conservative 58; Mismatches 162; Indels 94; Gaps 15;

QY 15 QEEPRYSYWCPC---KAEITPSSGKTLRLTYEENDESEVEIHTVSPMLT- 59
 DB 434 QKDSILYNGKWIISPDTFRFEITDRNSKILDVAVYDIPSSQMLTVANTSNLFGWIEKTK 493
 QY 60 -----PVLFTTR-----ADS-----FRYKXTGTANPKVTFKMSSEIMDAEGRI 98
 DB 494 DILSIPRPELKRMDYGIIDHADSRRGSHLFYF-TVFAKEPIQLTKGNWEVTGNGIYG 552

QY 99 VDEVRRLVYFEGTKDSPLHHLVYVSYV-----NPGVRLTLDRGSHSCCIS 146
 Db 553 YEVETDITFTFANBIGNMSQHLISLTDSTQNTQFOSLQNP-----SDYDPIDELS 606
 QY 147 QHCDFFISKYSNQNP-----HCVSLYKLSPPEDPTCKTKFEMATILD 190
 Db 607 SSARVAISKSLGPDTPIKVAGPLTRVLNVAEIHDDSLILQTLKDE-----KFKK----- 654
 QY 191 SAGLPDPTPELFEFESTT---GFTLYGMLYKPHDLOPKKPTTVLFIYGGPOVLVNN 247
 Db 655 ---KIKNYDLP-LTSYTWLMDGVELNYLIEKPMANIPKKYPIILVNIYGGPOVTF 710
 QY 248 RFKGVKFFRLNTLASLGYVVVVIDNRGSGHGLKFEAGFYKMGQIEIDQVE-GIQYLA 306
 Db 711 --KSLAFEGAVVSGLDVILQIEPRGTGKGMSFRMAEKLYGHEPRDITEVTKKFIQ 768
 QY 307 SRDPTDLDRVGHGMSYGYLSLML-MQRSDIFRVAIAGAPVTLMIFYDTGYT 360
 Db 769 RNSQHIDESKIAIWMSYGGFTSLKTVELNCGDTFKYAMAVALPVTMTLYDSVYT 823

RESULT 11

DPp6_BOVIN STANDARD; PRT; 863 AA.

ID DPp6_BOVIN STANDARD; PRT; 863 AA.
 AC P42659;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidyl)peptidase VI
 DE (Dipeptidyl)peptidase 6 (Dipeptidyl)peptidase IV like protein
 GN DPP6.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 NC NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=92108018; PubMed=1729689;
 RA Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family.";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=P42659-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=P42659-2; Sequence=VSP 005364;
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
 CC BRAIN, KIDNEY, OVARY AND TESTIS.
 CC -1- SIMILARITY: Belongs to peptidase family 89B.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M76428; AAC41622.1; -;
 CC EMBL; M76429; AAC41623.1; -;
 CC PIR; A41793; A41793.
 CC MEROPS; S09.973; -.

DR InterPro; IPR002469; DPPV N term.
 DR InterPro; IPR001375; Peptidase 59.
 DR InterPro; IPR000379; Ser esters site.
 DR Pfam; PF00930; DPPV N term; 1.
 DR Pfam; PF00326; Peptidase 59; 1.
 KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 FT DOMAIN 1 93
 FT TRANSMEM 94 114
 FT
 FT DOMAIN 115 863
 FT CARBOHYD 171 171
 FT CARBOHYD 402 402
 FT CARBOHYD 469 469
 FT CARBOHYD 533 533
 FT CARBOHYD 564 564
 FT CARBOHYD 811 811
 FT VARSPIC 1 79
 FT
 FT
 SQ SEQUENCE 863 AA; 96556 MW; 23DBA792B841A39D CRC64;
 Query Match 11.4%; Score 221; DB 1; Length 863;
 Best Local Similarity 21.0%; Pred. No. 4, Se-11;
 Matches 103; Conservative 58; Mismatches 169; Indels 160; Gaps 16;

QY 8 GVATFVLOAE-FDRYSGYWCCKAETTPSGKILRLIYENDESEVEIIVH--TSPLME 63
 Db 285 GSDMDLYEELIKTHIAHMSPDG-----TRLAVATINDSRVPELPIYTOSVVP 335
 QY 64 TRADSFRPKTGIANPKVTF-----KXSEIMI----- 91
 Db 336 T--AKRYHPKAGCENPISLSHVIGLNGPTHLEMTPPDDPRRREYITVWKVATSTKVA 393
 QY 92 -----DA-----EGR----- 96
 Db 394 VNMLSRAQVSLITLCDATITGVCTKKHEDSEAMILHRQNEBPVFSKDGKFFVRRAIPQG 453
 QY 97 -----IIVDEVRRLVYFEGTKDSPLHHL 121
 Db 454 GQGRFYHTVSSSQPNSSNDNIQSITSGMDVTKLISYBKRSQYIFLTEDLPRRQLY 513
 QY 122 VSYVNPGEVTR-----LTDRGYHSCCISGHCDFPISKYSNQNPCHVSLYKSSPED 175
 Db 514 SASVY--GSFNNQCLSCDLVNDCTVFSASFSGADEFLLKCBGPVP--TVSYVNTDKKK 570
 QY 176 DPTCKTKE-FMATILDSAGPLDYTPPELFEFESTTGTFLYGMLYKPHDLOFGKKYPTVL 234
 Db 571 MEDLETNEHVQKALSDROMPKVEYRKIE-----TDYKPLPIQLKPATFTDTAHYPLL 624
 QY 235 FIYGGPOVLVNNRFKGVKFFRLNTLASLGYVVVVIDNRGSGHGLKFEAGFYKMGQIE 294
 Db 625 VVDGTGPGQSVAEKF--AVTWETVWVSSHGAVVVCDDRGSGFQGTTRLHLVRRRLGSL 682
 QY 295 IDDOVEGQIYASRDPTDLDRVGHGMSYGYLSLMLMQRSD---IFRVAIAGAPV 350
 Db 683 EKDQNEAVRWVL--KEPYIDKTRVAVFGDYGGLSTYLLIPAKGDQAPVFSGSAISPIT 741
 QY 351 LMIFYDTGYT 360
 Db 742 DFKLYASAFS 751
 RESULT 12
 DPp6_RAT STANDARD; PRT; 859 AA.
 ID DPp6_RAT STANDARD; PRT; 859 AA.
 AC P46101;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidyl)peptidase VI
 DE (Dipeptidyl)peptidase 6 (Dipeptidyl)peptidase IV like protein
 GN DPP6.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 NC NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=92108018; PubMed=1729689;
 RA Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family.";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=P42659-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=P42659-2; Sequence=VSP 005364;
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
 CC BRAIN, KIDNEY, OVARY AND TESTIS.
 CC -1- SIMILARITY: Belongs to peptidase family 89B.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M76428; AAC41622.1; -;
 CC EMBL; M76429; AAC41623.1; -;
 CC PIR; A41793; A41793.
 CC MEROPS; S09.973; -.

GN DPP6.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN NCBI_TaxID=10116;
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
 RC TISSUE=Brain.
 RX MEDLINE=92108018; PubMed=1729689;
 RA Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family";
 CC Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
 CC -1- FUNCTION: May be involved in the physiological processes of brain
 CC function. Has no dipeptidyl aminopeptidase activity. The lack of
 CC activity may be due to the substitution of an aspartate residue
 CC for the serine residue in the proposed catalytic triad.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=46101-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=46101-2; Sequence=VSP 005366;
 CC -1- TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME
 CC PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS; IN
 CC CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M76426; AAC42061.1; -;
 DR EMBL: M76427; AAC42062.1; -;
 DR MEROPS: S09.973; -;
 DR InterPro: IPR002469; DPPX N term.
 DR InterPro: IPR001375; Peptidase S9.
 DR InterPro: IPR000379; Ser estersite.
 DR Pfam: PF000930; DPPX N term; 1.
 DR Pfam: PF00326; Peptidase S9; 1.
 KM Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 FT DOMAIN 1 89
 FT TRANSMEM 90 110
 FT SIGNAL-ANCHOR (POTENTIAL).
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 111 859
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 167 167
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 168 168
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 313 313
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 313 313
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 465 465
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 529 529
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 560 560
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 807 807
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 75
 FT MASLYORFTGKINTSSFPAPPEASHLGGQGEEDAGSKP
 FT LGPOADVAAPERGGAGRRFYOARSDCEDD -> MTT
 FT AKPSASGSKVQOQD (in isoform DPPX-S).
 FT /FTId=VSP 005366.
 SQ SEQUENCE 859 AA; 97301 MW; CE26856D26ED126B CRC64;
 Query Match 11.0%; Score 212; DB 1; Length 859;
 Best Local Similarity 19.7%; Pred. No. 2.7e-10;
 Matches 98; Conservative 67; Mismatches 158; Indels 174; Gaps 16;
 Oy 8 GAATPFLQRE-PDRYGVMMCPAETTPSGGKILRIYENDESEVFIHV---ISPMTE 63
 Db 281 GSDMLYEEELIKSHIAHWSPDG-----TRLAYATINDSRVPLMELPYTGSVYP 331
 Oy 64 TRRADSFYPKTGTANPKVTF-----KMSBIMI----- 91

Db 332 T--VKPYHYPKAGSENPISLHVLGNQPTHDLFEMPPDDPRMEBYITVWKMASTKVA 389
 Oy 92 -----DA-----BGR----- 96
 Db 390 VTWLNRAQNSILTLCDATGTGCTKHEDESEAWLHRONEBPVPSKDRKFFVRAIPQG 449
 Oy 97 -----IIVDEVRRLVYEGTDSPLEHLY 121
 Db 450 GRGKFHITVSSSPNSNDNISITSGDMVTEILTYDEKANKLYFLSTEDLPRRRALY 509
 Oy 122 VSYVNDGEVTRLTDRGYSHSCCISQHCDFPISKYSQKNPCVSLYKLSPP-----ED 175
 Db 510 SANTVDDENRCL-----SCDLVENCTVYSAFSHMD--FFLKCEBPVPTVYH 559
 Oy 176 DPTCKTEP-----WATLDSAGLPDTPPEIFTSFSTGTGLYGMKYKPHDLPQ 227
 Db 560 NTDKRMFPDLEANEQOKALYDQMKRKEIRKIEV-----EDYSLEPMQILKATPDT 613
 Oy 228 KKYPTVLFIVGPOVLNRRFKGVYFRLNTLASLGVVVYVINDRGSCHRLKFEQAFK 287
 Db 614 AHYPLLVDNQTSGSQSVSRFEVT--WEYLVSSHQAIVVYKCDGSGFGGTYLHEVR 671
 Oy 288 YKMGQIHDQVEGLQYLAIRYFDLDRVGIHGSYGYLSLMALQKSD----IFRVA 343
 Db 622 RRLGQLERKQMEAVRMTL-KEGYIDKTRVAVFQKQGYSTYILPAKGENQGTFTCG 730
 Oy 344 IAGAPVLMTFYDNGY 360
 Db 731 SALSPITDFLYASAFS 747
 RESULT 13
 DPP6 MOUSE STANDARD; PRT; 804 AA.
 ID DPP6 MOUSE Q92218; Q92219;
 AC Q92218; Q92219; Q92219;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
 GN DPP6 OR DPP-6.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryo;
 RX MEDLINE=99030650; PubMed=9811881;
 RA Hough R.B., Lenggeling A., Bedian V., Lo C., Bucan M.;
 RT "Pump white inversion in the mouse disrupts dipeptidyl aminopeptidase-
 RT like protein 6 and causes dysregulation of kit expression.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:13800-13805(1998).
 CC -1- FUNCTION: May be involved in the physiological processes of brain
 CC function. Has no dipeptidyl aminopeptidase activity. The lack of
 CC activity may be due to the substitution of an aspartate residue
 CC for the serine residue in the proposed catalytic triad (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF092507; AAC97366.1; -;
 DR EMBL: AF092506; AAC97365.1; -;

EMBL: AF092505; AAC98381.1; -
 DR MEROPS; S09_973; -
 DR MGD; MGI:94921; DPP6.
 DR InterPro; IPR002469; DPPV N term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR00379; Ser esterase site.
 DR Pfam; PF00930; DPPV N term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR Transmembrane; Glycoprotein; Signal-anchor.
 KW DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 35 55 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 56 804 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 440 455 MISSING (IN REF. 1; AAC98381).
 FT CONFLICT 638 638 Q -> P (IN REF. 1; AAC97365).
 SQ SEQUENCE 804 AA; 91260 MW; 09CFE7AD8A7168 CRC64;

Query Match 10.9%; Score 211; DB 1; Length 804;
 Best Local Similarity 19.7%; Pred. No. 3e-10;
 Matches 98; Conservative 65; Mismatches 159; Indels 176; Gaps 16;

QY 8 GVATFVLQEE-PRYSGYMPCPAETTPSGGKILRIEYENDESEVEIHHV---TSPMLE 63
 DB 226 GLSDWLYEEELIKSLAHMWSPDG-----TRLATYATINDSRVPLMELPHYTSVVP 276
 QY 64 TRADSPRYKKTGANPKVTF-----KXSEIMI----- 91
 DB 277 T--VKRYHPKASSENSISLHVIGLNGPHTLEMPDDPRKREYIITVWKATSTKVA 334
 QY 92 -----DA-----EGR----- 96
 DB 335 VFWLNRAQNVSLITLQDATTCVCTKKHDESEAMIRQNEPEFSGDKGRFFVRALPDG 394
 QY 97 -----IIIVDERLVYFEETKDSPLRHLY 121
 DB 395 GRGKFYHITVSSSQPNSSNDNIQSITSGDWDTKILSYDEKRMKIFLSTEDLPRRHLY 454
 QY 122 VVSYNVGEVTRITLDGYSHSCICSHODPFISKNSQKPHCVSLYKLSPEDDPTC-- 179
 DB 455 SANTVDPNRQCL-----SCDLVENCYVSASFHNDM--FLLKCEGP-GVPLTVV 503
 QY 180 -----KTRFMTATLDSAGPLPDYTPPEIFSFESTGTFLYGLMYKPPHLDQ 226
 DB 504 HNTTDKRMEDLEANEVEQKAINDRQMKIEYKIEV-----EDYSLPMQILKPAFTFD 557
 QY 227 GKRYFVFLTYIGSPQQLVNNRKYKYPFLNLTALSLGYVVVITDNGSCHRLKEGAF 286
 DB 558 TAHYPLLVLDGTPGQSVTEREVEVT--WETVIVSHGAVVVCDDRGSGFGQTKLQEV 615
 QY 287 KYMGOIISIDDQVEGLOYLASRYDFTLDRVGTHGWSYGGYSLMALMQRSD----IFRY 342
 DB 616 RRRIGLEBKQDGEAVRITML-KQYIDKTRAVAVFGDYGISTYITLPAKEMOGQTFTC 674
 QY 343 AIAGAPVTLMIIFYDTYT 360
 DB 675 GSALSPITDPFKLYASAFS 692

RESULT 14

DPPE HUMAN STANDARD; PRT; 865 AA.
 AC P42658;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
 OS DPP6.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
 RC TISSUE=Hippocampus;
 RX MEDLINE=93372805; Pubmed=8103397;
 RA Yokotani N., Doi K., Wenthold R.J., Wada K.;
 RT "Non-conservation of a catalytic residue in a dipeptidyl
 aminopeptidase IV-related protein encoded by a gene on human
 chromosome 7.";
 RL Hum. Mol. Genet. 2:1037-1039(1993).
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
 ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
 FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=P42658-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=P42658-2; Sequence=VSP 005365;
 CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
 CC -!- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M96859; AAA35760.1; -
 DR EMBL; M96860; AAA35761.1; -
 DR PIR; I54331; I54331.
 DR MEROPS; S09_973; -
 DR GeneW; HANC:3010; DPP6.
 DR MIM; 126141; -
 DR GO; GO:0008239; F.dipeptidyl-peptidase activity; TAS.
 DR InterPro; IPR002469; DPPV N term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR00379; Ser esterase site.
 DR Pfam; PF00930; DPPV N term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 KW DOMAIN 1 95 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 96 116 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 117 865 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 813 813 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
 MASLQRFYFGKINTSRSPAPPEASHLLGGQEPEDGAGAG
 KPLGPRQAQAARBERGGGAGGAGRRPFGYQGSDDEED
 -> MTTAKEPSASGKSVQQQOE (in isoform
 DPPX-S).
 FT FT /FTId=VSP 005365.
 SQ SEQUENCE 865 AA; 97588 MW; 14B1ABE0024464B CRC64;

Query Match 10.5%; Score 203; DB 1; Length 865;
 Best Local Similarity 19.6%; Pred. No. 1.6e-09;
 Matches 98; Conservative 65; Mismatches 157; Indels 180; Gaps 17;


```

QY 8 GVATFVLQEE-FDRSYGWMCPKAEFTPSGGKILRLIYEENDESEFVRIHV---TSPMLE 63
DB 287 GLSDMLVEBELIKTHIAMWSPD-----TRIAIAINDSVIPIMELPTTGTSTYR 337
QY 64 TRRADSRYPKPTGTANPKVTF-----KMSIIMI- 91
DB 338 T--VKPYHYPKAGSENPSISLHVIGLNGPTHDLBMMPPDDPRMREYITVMKMASTSKVA 395
QY 92 -----Da-----ER----- 96
DB 366 VTMLNRAQNVSLILLCDAFTGVCYTRKHDESEAMLHRONEEPVSKDGRKFFTRAIPQG 455
QY 97 -----IIVDEVRLIYVEFGTQSPLEHTLY 121
DB 456 GNGKPYHITVSSSQNSNDNIQSTISGDMVTKLALADENGNKIYFLSTEDLRRRLQY 515
QY 122 VVSIVNPGEVTRILTRGISH---SCCISQCHDFPISKYSNOKNPHCVSLYLSBP----- 173
DB 516 SAN-----TEGNFRROCLSCDLVENCITYFSASFSSHMD--FELLCEGPGVPMV 562
QY 174 -EDDPCTKTEF-----WATLDSAGPLDPYPRPIFSFESTGTLYGMLYKPHDL 224
DB 563 TVHNTTDKKMFDELTEHNVKKAINDROMPKVEYDIDI-----DYNLMQILKRPATF 616
QY 225 QPGKKYPTVLFYGGPQVQVYNNRPKVKYFRLLNTLASLGYVVVINDRSGCHGLKEEG 284
DB 617 TDTTHYPLDLVVDGTPGSSQVAKEFE--VSWETVWVSSHGAIVVVCDDRGSGGPGTKLH 674
QY 285 AKYKMGQIEIDDOYEGIOYLAIRDYFDLDRVGHGWSYGYSLMALMQRSD---IF 340
DB 675 EVRRRLGLEEKQDQHEAFTML-KEQYIDRTKVAVFGVDGYGYSTYLTPAKGENQGT 733
QY 341 RVALIAGAPVTLWIFDYTGTY 360
DB 734 TCGSALSPITDFKLYASAFS 753

RESULT 15
YDZF SCHPO STANDARD; PRT; 853 AA.
AC Q9P7E9; 013720;
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative dipeptidyl aminopeptidase C14C4.15c (EC 3.4.14.-).
GN SPAC14C4.15C OR SPAP760.01C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxId=4896;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels J.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovich E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Wolckaert G., Aert R., Robben J., Grymoprez B.,
RA Welfens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Moestl D., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

```

```

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Sipakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.",
RL Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
CC vacuoles.
CC -1- SIMILARITY: Belongs to peptidase family 99B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: A1162631; CAB81084.1; -.
DR EMBL: Z98596; CAB1208.1; -.
DR GeneDB: SPombe; SPAC14C4.15c; -.
DR InterPro: IPR002469; DPIP_V_N-term.
DR InterPro: IPR001375; Peptidase_S9.
DR Pfam: PF00930; DPIP_V_N-term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
KM Hypothetical protein, Hydrolase, Aminopeptidase, Dipeptidase;
KM Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 66 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 67 89 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 90 853 LUMENAL (POTENTIAL).
FT ACT_SITE 719 719 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 795 795 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 828 828 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 853 AA; 98341 MW; 38450BA50F8304B6 CRC64;

Query Match 10.2%; Score 197.5; DB 1; Length 853;
Best Local Similarity 29.8%; Pred. No. 4.7e-09;
Matches 53; Conservative 25; Mismatches 59; Indels 41; Gaps 6;

QY 200 PPEIFSESTTGTYLGMLYKPHDLQPGKKYPTVLFYGGPQVQVYNNRPKVKYFRLLNT 259
DB 602 PSVFFKVKIKVNTAIIQEIIRPNFNRKRYPTVFLYGAPOSALVTGYE---MDINE 657
QY 260 LASLGYVVVINDRSGCHRLKPEGAFFKMGQIEIDDOVEG-----LQY 304
DB 658 LMASVNFVLT-----KVDIVLSD--VSGQHLPFSDHELLIKSWML 698
QY 305 LASRYD-FITLDRVGHGWSYGYSLMALMQRSDIFPVVAIAGAPVTLWIFDYTGTY 360
DB 699 LRSYVDTPTIDRHRVGHGWSFGGYLT-LKIENADPIRTGAVVAVPTDMRYDAYYS 755

Search completed: October 15, 2003, 17:11:34
Job time : 7.71145 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:02:16 ; Search time 23.0243 Seconds

(Without alignments)
4034,822 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1933
Sequence: 1 EDARSAGVATFVLQSEFDR.....RVAIAGAPVTLMIFYDTGYT 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 segs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP Vertebrate:*
14: SP Unclassified:*
15: SP_virus:*
16: SP_bacteriap:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	1933	100.0	360	4 Q9HBM2	Q9HBM2 homo sapien
2	1836.5	95.0	882	4 Q9HBM5	Q9HBM5 homo sapien
3	1836.5	95.0	882	4 Q81WG7	Q81WG7 homo sapien
4	1756.5	90.9	892	11 Q9D4G6	Q9D4G6 mus musculu
5	1534	79.4	831	4 Q8NEM5	Q8NEM5 homo sapien
6	1528	79.0	832	4 Q96JX1	Q96JX1 homo sapien
7	1220.5	63.1	628	4 Q8N3F5	Q8N3F5 homo sapien
8	1220.5	63.1	863	4 Q8W4X8	Q8W4X8 homo sapien
9	1204.5	62.3	862	11 Q8BWT9	Q8BWT9 mus musculu
10	1204.5	62.3	862	11 Q8BVG4	Q8BVG4 mus musculu
11	1086.5	56.2	310	4 Q9HBM4	Q9HBM4 homo sapien
12	984	50.9	439	4 Q9HVR3	Q9HVR3 homo sapien
13	944	48.8	508	4 Q75273	Q75273 homo sapien
14	860.5	44.5	465	4 Q9HBM3	Q9HBM3 homo sapien
15	799	41.3	312	4 Q96NT8	Q96NT8 homo sapien
16	722	37.4	1042	5 Q9VC20	Q9VC20 drosophila

17	722	37.4	1102	5 Q9VC19	Q9VC19 drosophila
18	718.5	37.2	1053	5 Q81H07	Q81H07 drosophila
19	527.5	27.3	738	16 Q9AGE0	Q9AGE0 caulobacter
20	522	27.0	469	4 Q9NKR4	Q9NKR4 homo sapien
21	494	25.6	763	16 Q8EAB7	Q8EAB7 shewanella
22	491.5	25.4	927	5 Q965K3	Q965K3 caenorhabdi
23	491.5	25.4	931	5 Q498B7	Q498B7 caenorhabdi
24	471	24.4	751	16 Q8P3V8	Q8P3V8 xanthomonas
25	463	24.0	757	16 Q8PPD7	Q8PPD7 xanthomonas
26	430.5	22.3	741	2 P95782	P95782 xanthomonas
27	422.5	21.9	746	10 Q9ENR6	Q9ENR6 arabidopsis
28	390.5	20.2	552	10 Q8GJ07	Q8GJ07 arabidopsis
29	369	19.1	711	2 Q47900	Q47900 flavobacter
30	366.5	19.0	730	2 Q93JY4	Q93JY4 prevotella
31	347.5	18.0	771	3 Q42812	Q42812 aspergillus
32	346	17.9	723	2 Q31048	Q31048 porphyromon
33	344	17.8	723	2 Q66223	Q66223 porphyromon
34	343.5	17.8	432	4 Q75868	Q75868 homo sapien
35	322	16.7	237	2 Q87543	Q87543 capnocytoph
36	320	16.6	765	3 Q14425	Q14425 aspergillus
37	319	16.5	827	16 Q8PPU4	Q8PPU4 xanthomonas
38	317.5	16.4	748	13 P70092	P70092 xenopus lae
39	313.5	16.2	901	3 Q96VT7	Q96VT7 streptococ
40	304.5	15.8	707	16 Q9P348	Q9P348 streptococ
41	291	15.1	803	16 Q8P500	Q8P500 xanthomonas
42	288	14.9	167	4 Q8N2J7	Q8N2J7 homo sapien
43	282	14.6	935	5 Q9VMB4	Q9VMB4 drosophila
44	281	14.5	745	5 Q9VMM2	Q9VMM2 drosophila
45	271	14.0	802	5 Q9VUH1	Q9VUH1 drosophila

ALIGNMENTS

RESULT 1

Q9HBM2 ID Q9HBM2 PRELIMINARY; PRT; 360 AA.
AC Q9HBM2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Dipeptidyl peptidase 8 (Fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollett E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL; AF221637; AAG29769.1; -
DR MEROPS; S09.018; -
DR InterPro; IPR01375; Peptidase_S9.
DR InterPro; IPR00379; Set_struct_site.
DR Pfam; PF00326; Peptidase_S9; 1.
FT NON_TER 1
FT TER 360
SQ SEQUENCE 360 AA; 41070 MW; CF81COBB61423E22 CRC64;

Query Match 100.0%; Score 1933; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 3,7e-168;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDARSAGVATFVLQSEFDRSGYWCCKATETTS9GKILILYENDESSEVETHTVSP 60
DB 1 EDARSAGVATFVLQSEFDRSGYWCCKATETTS9GKILILYENDESSEVETHTVSP 60
QY 61 MLETRADSFYRPTGTAMPKVTTFKMSHIMIDAGRIIVDVRRLVYEGTKDSPLEHNL 120

```

Db      61 MLETRADSFRRYKGTANPKVTFKMSIIMDAEGRIIVDEVRRLVYEGKIDSLHHL 120
QY      121 YVSVYVNGEYRLTRDGRYSHSCCISQCHDFPISKYSNQKNPHCVSLYKLSPEDDPTCK 180
Db      121 YVSVYVNGEYRLTRDGRYSHSCCISQCHDFPISKYSNQKNPHCVSLYKLSPEDDPTCK 180
QY      181 TKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGP 240
Db      181 TKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGP 240
QY      241 QVOLVNNRFKGVKRYFRNLNTLASLGVVVVVIDNRGSCHRGLKFEKAFKYMGOIEIDDOVE 300
Db      241 QVOLVNNRFKGVKRYFRNLNTLASLGVVVVVIDNRGSCHRGLKFEKAFKYMGOIEIDDOVE 300
QY      301 GLQYIASRYDFIDLRVGIHGMVSGYGLSLMALMQRSDFRVALAGAPVTLMIFYDTGYT 360
Db      301 GLQYIASRYDFIDLRVGIHGMVSGYGLSLMALMQRSDFRVALAGAPVTLMIFYDTGYT 360

```

RESULT 2

```

ID      Q8HBM5      PRELIMINARY;      PRT;      882 AA.
AC      Q8HBM5;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT      01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE      Dipeptidyl peptidase 8.
GN      DPP8.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_Taxid=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20467194; PubMed=11012666;
RA      Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
RA      Gorrell M.D.;
RT      "Cloning, expression and chromosomal localization of a novel human
RT      dipeptidyl peptidase (DPP) IV homolog, DPP8."
RL      Eur. J. Biochem. 267:6140-6150(2000).
DR      EMBL: AF221634; AAC29766.1; -.
DR      MEROB8; S09_018; -.
DR      InterPro: IPR002469; DPPIV_N_term.
DR      InterPro: IPR001375; Peptidase_S9.
DR      InterPro: IPR000379; Ser esters_site.
DR      Pfam: PF00930; DPPIV_N_term; 1.
DR      Pfam: PF00326; Peptidase_S9; 1.
SQ      SEQUENCE 882 AA; 101421 MW; AD80IC302DB4652B CRC64;

```

```

Query Match      95.0%; Score 1836.5; DB 4; Length 882;
Best Local Similarity 67.5%; Pred. No. 8.4e-159;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

```

```

QY      1 EEDARAGVATFVLQEFPRYSGYVWCPKAEITPSSGKILRIIYEENDESEVEIIHVTSP 60
Db      244 EEDARAGVATFVLQEFPRYSGYVWCPKAEITPSSGKILRIIYEENDESEVEIIHVTSP 303
QY      61 MLETRADSFRRYKGTANPKVTFKMSIIMDAEGRII----- 98
Db      304 MLETRADSFRRYKGTANPKVTFKMSIIMDAEGRIIIVIDKELIQPEILFEGVEYIA 363
QY      99 ----- 98
Db      364 RAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERQRLIESVPSVTPLIY 423
QY      99 ----- 98
Db      424 EETTDIWINIHDIHFVFPQSHHEIEIFIASCKTGFRHLKYITSLIKSKYKRSSGGLP 483
QY      99 -----VDEVRRLVYFEGTKDPSLEHLLYVSVYN 127
Db      484 APSDFKCPIKEELIATSGEWEVLGRHGSNIQDEVRRLVYFEGTKDPSLEHLLYVSVYN 543

```

RESULT 3

```

ID      Q8IMG7      PRELIMINARY;      PRT;      882 AA.
AC      Q8IMG7;
DT      01-MAR-2003 (Tremblrel. 23, Created)
DT      01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT      01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE      Similar to dipeptidyl peptidase 8.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_Taxid=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Straussberg R.;
RL      Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
DR      EMBL: BC040203; AAH40203.1; -.
SQ      SEQUENCE 882 AA; 101391 MW; 88C76A5BCE707F9 CRC64;

```

```

Query Match      95.0%; Score 1836.5; DB 4; Length 882;
Best Local Similarity 67.5%; Pred. No. 8.4e-159;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

```

```

QY      1 EEDARAGVATFVLQEFPRYSGYVWCPKAEITPSSGKILRIIYEENDESEVEIIHVTSP 60
Db      244 EEDARAGVATFVLQEFPRYSGYVWCPKAEITPSSGKILRIIYEENDESEVEIIHVTSP 303
QY      61 MLETRADSFRRYKGTANPKVTFKMSIIMDAEGRII----- 98
Db      304 MLETRADSFRRYKGTANPKVTFKMSIIMDAEGRIIIVIDKELIQPEILFEGVEYIA 363
QY      99 ----- 98
Db      364 RAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERQRLIESVPSVTPLIY 423
QY      99 ----- 98
Db      424 EETTDIWINIHDIHFVFPQSHHEIEIFIASCKTGFRHLKYITSLIKSKYKRSSGGLP 483
QY      99 -----VDEVRRLVYFEGTKDPSLEHLLYVSVYN 127
Db      484 APSDFKCPIKEELIATSGEWEVLGRHGSNIQDEVRRLVYFEGTKDPSLEHLLYVSVYN 543
QY      128 PGEVTRLRDGRYSHSCCISQCHDFPISKYSNQKNPHCVSLYKLSPEDDPTCKTKEFWAT 187
Db      544 PGEVTRLRDGRYSHSCCISQCHDFPISKYSNQKNPHCVSLYKLSPEDDPTCKTKEFWAT 603
QY      188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGPOVOLVNN 247
Db      604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGPOVOLVNN 663
QY      188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGPOVOLVNN 247
Db      604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIOPGKKYPTVLFIYGGPOVOLVNN 663
QY      248 RFGVKYFRNLNTLASLGVVVVVIDNRGSCHRGLKFEKAFKYMGOIEIDDOVEGLQYIAS 307
Db      724 RYDFIDLRVGIHGMVSGYGLSLMALMQRSDFRVALAGAPVTLMIFYDTGYT 776
QY      308 RYDFIDLRVGIHGMVSGYGLSLMALMQRSDFRVALAGAPVTLMIFYDTGYT 360
Db      724 RYDFIDLRVGIHGMVSGYGLSLMALMQRSDFRVALAGAPVTLMIFYDTGYT 776

```

QY 308 RYDFIDLRVGIHMSYGGYLSLMAIMORSDFRVAIAGAPVTLWIFDYDTGYT 360
 DB 724 RYDFIDLRVGIHMSYGGYLSLMAIMORSDFRVAIAGAPVTLWIFDYDTGYT 776

RESULT 4

QD4G6 PRELIMINARY; PRT; 892 AA.

ID QD4G6 PRELIMINARY; PRT; 892 AA.

AC QD4G6 PRELIMINARY; PRT; 892 AA.

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Adult male testis cDNA, RIKEN full-length enriched library,
 clone:493434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,
 Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 Kawachi Y., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
 Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numasaki R., Ono M.,
 Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
 Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 Tejima Y., Toyota T., Yamamuro T., Yasunishi A., Yoshida K., Yoshino M.,
 Yamamatsu M., Hayashizaki Y.,
 Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA The PANTOM Consortium,
 the RIKEN Genome Exploration Research Group Phase I & II Team;
 "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs";
 Nature 420:563-573 (2002).

RL [3]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA MEDLINE=21085660; PubMed=11217851;
 RIKEN PANTOM Consortium;
 "Functional annotation of a full-length mouse cDNA collection";
 Nature 409:685-690 (2001).

RL [4]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA MEDLINE=99279253; PubMed=10349636;
 Carninci P., Hayashizaki Y.;
 "High-efficiency full-length cDNA cloning";
 Meth. Enzymol. 303:19-44 (1999).

RL [5]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA MEDLINE=20499374; PubMed=11042159;
 Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes";
 Genome Res. 10:1617-1630 (2000).

RL [6]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA MEDLINE=20530913; PubMed=11076861;
 Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 Kono H., Akiyama J., Nishi K., Kitsumi T., Tashiro H., Itoh M.,
 Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa Y., Tanaka T., Matsura S., Kawai J.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multichipillary sequencer";
 RL Genome Res. 10:1757-1771 (2000).

DR EMBL; AK016546; BAB30295.2; -
 SO SEQUENCE 892 AA; 102284 MW; PDE0DBEDCA4CA346 CRC64;

Query Match 90.9%; Score 1756.5; DB 11; Length 892;
 Best Local Similarity 64.4%; Pred. No. 1.7e-151;
 Matches 343; Conservative 7; Mismatches 10; Indels 173; Gaps 1;

QY 1 EEDARAGATATVLOEPEFRVSGYVWCPKAEFTPSGCKTIRLYEENDESEVEIHHVSP 60
 DB 254 EEDPRAGATATVLOEPEFRVSGYVWCPKAEFTPSGCKTIRLYEENDESEVEIHHVSP 313
 QY 61 MLETRADSPFRPKGTANPKVTFKMSIEMDAEGR11----- 98
 DB 314 MLETRADSPFRPKGTANPKVTFKMSIEMDAEGR11----- 98
 QY 99 ----- 98
 DB 374 RAGWTPGKHAMSLIDRSQTHLQVILSPDLFPEVDAMDORLIESVDPSTPLIY 433
 QY 99 ----- 98
 DB 434 EETTDWINIHDFHVFPOTHEDLEFIASCKTGFRHLKYITSLIKSKYKRSSGGLP 493
 QY 99 ----- 127
 DB 494 APSDFKCPKEBIIITSGEWEVLGRHNSIWDEARKVYFPGTDSPLEHLLVYTSVAN 553
 QY 128 PGEVTRLPDRGSHSCISQHCDFISKYSKONCHCSLYLSPEDDPCTKEPFNAT 187
 DB 554 PGEVTRLPDRGSHSCISQHCDFISKYSKONCHCSLYLSPEDDPCTKEPFNAT 613
 QY 188 IIDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLPQKKYPTVLFYGGPOVQLVNN 247
 DB 614 IIDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLPQKKYPTVLFYGGPOVQLVNN 673
 QY 248 REFQVYFRPLNTLASLGYVVVVYIDNRGSHRDLKREGAFKYMGGIETIDDOVEGIQLIAS 307
 DB 674 REFQVYFRPLNTLASLGYVVVVYIDNRGSHRDLKREGAFKYMGGIETIDDOVEGIQLIAS 733
 QY 308 RYDFIDLRVGIHMSYGGYLSLMAIMORSDFRVAIAGAPVTLWIFDYDTGYT 360
 DB 734 RYDFIDLRVGIHMSYGGYLSLMAIMORSDFRVAIAGAPVTLWIFDYDTGYT 786

RESULT 5

Q8NEMS PRELIMINARY; PRT; 831 AA.

ID Q8NEMS PRELIMINARY; PRT; 831 AA.

AC Q8NEMS PRELIMINARY; PRT; 831 AA.

DT 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DE Similar to dipeptidylpeptidase 8.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC030688; AAH30688.1; -
 SO SEQUENCE 831 AA; 95527 MW; 0B2A1A2FE70CB22 CRC64;

Query Match 79.4%; Score 1534; DB 4; Length 831;
 Best Local Similarity 58.0%; Pred. No. 3.4e-131;

Matches 309; Conservative 0; Mismatches 0; Indels 224; Gaps 2;

QY 1 EDARSAGVATFVLOEEDFRYSGYWCPCRAETTPSGKILRLIYENDESEVEIITHVSP 60
 DB 244 EDARSAGVATFVLOEEDFRYSGYWCPCRAETTPSGKILRLIYENDESEVEIITHVSP 303
 QY 61 MLETRRADSPFRYPKTGTANPKVTFKMSIIMIDAGR11-----98
 DB 304 MLETRRADSPFRYPKTGTANPKVTFKMSIIMIDAGR11DVIDKELIQFEILLFEGVEYIA 363
 QY 99 -----98
 DB 364 RAGWTPBKGKAMSILLDRSQTRLQIVLISPFLIPVEDDVMERQRLIESVDSVTPLIITY 423
 QY 99 -----98
 DB 424 EETDWINIHDIHFVPPQSHSEIEIFASECKTGFRHLKYITSILKSKYKSSGGLP 483
 QY 99 -----VDEVRLVYFEGTKDSPLEHLLVVSIVN 127
 DB 484 APSDFKPIKEEIAITSGEMEVLRHGSNIQVDEVRLVYFEGTKDSPLEHLLVVSIVN 543
 QY 128 PGEVTRLTDRGYSHSCCISQHCDFPISKYSNQKNPHCVSLYKLSPPDDPTCKTKEFWAT 187
 DB 544 PGEVTRLTDRGYSHSCCISQHCDFPISKYSNQKNPHCVSLYKLSPPDDPTCKTKEFWAT 603
 QY 188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIQPKKYPVLFYIGGPQVQLVNN 247
 DB 604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIQPKKYPVLFYIGGPQVQLVNN 663
 QY 248 RFKGVKFRINTLASLGYVVVVIDNRGSCHRGLKFEAGFYKMQQIIEIDQVEBQLYIAS 307
 DB 664 RFKGVKFRINTLASLGYVVVVIDNRGSCHRGLKFEAGFYKMQQIIEIDQVEBQLYIAS 706
 QY 308 RYDFIDLDVGIHMSYGYLSLMLMQRSDFRVALAGAPVTLMIFYDTGYT 360
 DB 707 -----VALAGAPVTLMIFYDTGYT 725

RESULT 6
 Q96JX1 PRELIMINARY; PRT; 632 AA.
 AC Q96JX1;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DB Hypothetical protein FLJ14920.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
 RA Nimomiya K., Iwayanagi T.;
 RT "NEBO human cDNA sequencing project."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK027826; BAB55395.1; -;
 DR MEROPS; S09.018; -;
 KW Hypothetical protein.
 SQ SEQUENCE 632 AA; 72639 MW; 9BDF598B06985AA4 CRC64;

Query Match 79.0%; Score 1528; DB 4; Length 632;
 Best Local Similarity 57.8%; Pred. No. 8.2e-131;
 Matches 308; Conservative 0; Mismatches 1; Indels 224; Gaps 2;

QY 1 EDARSAGVATFVLOEEDFRYSGYWCPCRAETTPSGKILRLIYENDESEVEIITHVSP 60

DB 45 EDARSAGVATFVLOEEDFRYSGYWCPCRAETTPSGKILRLIYENDESEVEIITHVSP 104
 QY 61 MLETRRADSPFRYPKTGTANPKVTFKMSIIMIDAGR11-----98
 DB 105 MLETRRADSPFRYPKTGTANPKVTFKMSIIMIDAGR11DVIDKELIQFEILLFEGVEYIA 164
 QY 99 -----98
 DB 165 RAGWTPBKGKAMSILLDRSQTRLQIVLISPFLIPVEDDVMERQRLIESVDSVTPLIITY 224
 QY 99 -----98
 DB 225 EETDWINIHDIHFVPPQSHSEIEIFASECKTGFRHLKYITSILKSKYKSSGGLP 284
 QY 99 -----VDEVRLVYFEGTKDSPLEHLLVVSIVN 127
 DB 285 APSDFKPIKEEIAITSGEMEVLRHGSNIQVDEVRLVYFEGTKDSPLEHLLVVSIVN 344
 QY 128 PGEVTRLTDRGYSHSCCISQHCDFPISKYSNQKNPHCVSLYKLSPPDDPTCKTKEFWAT 187
 DB 345 PGEVTRLTDRGYSHSCCISQHCDFPISKYSNQKNPHCVSLYKLSPPDDPTCKTKEFWAT 404
 QY 188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIQPKKYPVLFYIGGPQVQLVNN 247
 DB 405 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIQPKKYPVLFYIGGPQVQLVNN 464
 QY 248 RFKGVKFRINTLASLGYVVVVIDNRGSCHRGLKFEAGFYKMQQIIEIDQVEBQLYIAS 307
 DB 465 RFKGVKFRINTLASLGYVVVVIDNRGSCHRGLKFEAGFYKMQQIIEIDQVEBQLYIAS 507
 QY 308 RYDFIDLDVGIHMSYGYLSLMLMQRSDFRVALAGAPVTLMIFYDTGYT 360
 DB 508 -----VALAGAPVTLMIFYDTGYT 526

RESULT 7
 Q8N3F5 PRELIMINARY; PRT; 628 AA.
 AC Q8N3F5;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DB Hypothetical protein (fragment).
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Melanoma;
 RA Ansoorge W., Winkner U., Nemes H.W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL834376; CND39039.1; -;
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF00326; Peptidase_S9; 1.
 KW Hypothetical protein.
 FT NON TER 1
 SQ SEQUENCE 628 AA; 71368 MW; DCOB6A9440507C3B CRC64;

Query Match 63.1%; Score 1220.5; DB 4; Length 628;
 Best Local Similarity 44.2%; Pred. No. 1e-102;
 Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;

QY 2 EDARSAGVATFVLOEEDFRYSGYWCPCRAETTPSGG-KILRLIYENDESEVEIITHVSP 60
 DB 26 DDPKSHGAVTFVLOEEDFRYSGYWCPCRAETTPSGGKILRLIYENDESEVEIITHVSP 85
 QY 61 MLETRRADSPFRYPKTGTANPKVTFKMSIIMIDAGR11-----98
 DB 86 ALFERKTDSPFRYPKGTANPKVTFKMSIIMIDAGR11DVIDKELIQFEILLFEGVEYIA 145

```

QY 99 ----- 98
Db 146 RAGWTRDGKYAMAMFLDRPQOMLOLVLLPPALFIPSTENEBORLASARAVPRNVQPYVY 205
QY 99 ----- 98
Db 206 EEVTNWIMVNDIFYPPNQRBERTSAFAPMNAAPASAICTKSPPLKSGDYWSEPPS 265
QY 99 ----- 98
Db 266 PGEDEFKCPKEBIALTSGMEVILARHSGKIWNNEETKLVYFQGTQDTPLEHLLYVVSYE 325
QY 127 NPGEVTRLTDGYSGCCISQHCDFIISKYNQKNPHCVSLYKLSPEDDPTCKTEFWA 186
Db 326 AAGGIIVLTPRGFSHSCMSQNFDMFVSHSSYSTPCVHYVLTSGPDDDLPHKQRFWA 385
QY 187 TLIDSAGPLPDYTPPEIIFSPFSTGTFTLYGMLYKPHDLPQKKYPTVLFIYGPQVQLVN 246
Db 386 SMMEASCPDYPVPEIIFHFTSRDVLXGMLYKPHALQPKKHPVLFIYGGPQVQLVN 445
QY 247 NREKGVYFRLNTLASIGYVVVVIDNNGSCHRGKFEKGAFFYKMGQIEIDQVEGLQVLA 306
Db 446 NSFEGIKYRLNTLASIGYVVVVIDNNGSCHRGKFEKGAFFYKMGQIEIDQVEGLQVLA 505
QY 307 SRYDFIDLDVYGIGMSYGYLSLALMQRSDIFRYALAGAPYTLFIYDGYT 360
Db 506 EKXGFIIDSRVAIHGMSYGYLSLALMGLIHKQVFKVALAGAPYTMAYDGYT 559

RESULT 8
Q8WXD8 PRELIMINARY; PRT; 863 AA.
ID Q8WXD8
AC Q8WXD8;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DB Dipeptidyl peptidase 9.
GN DP9.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Olsen C.; Wegmann N.;
RT "Identification and characterization of a novel member of the
RT dipeptidyl peptidase IV-related family.";
RL Submitted (NCV-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Tissue=Skin;
RC Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF452102; AL47179.1; -.
DR EMBL; BC037948; AAH37948.1; -.
DR MEROPS; S09.019; -.
DR InterPro; IPR002469; DDPV N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DDPV N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 863 AA; 98263 MW; 40FE0B78E26CDED5 CRC64;

Query Match 63.1%; Score 1220.5; DB 4; Length 863;
Best Local Similarity 44.2%; Pred. No. 1.6e-102;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;

QY 2 EDARSAGVATFVLOEFPDRYSGYWCPCAKETTPSGG-KILRIYEENDESVEIIVTSP 60
Db 234 DPKSAGVATFVLOEFPDRYSGYWCPCAKETTPSGG-KILRIYEENDESVEIIVTSP 293
QY 61 MLETRADSPRYPKTGTANPKVTFKMSIIMIDABGRIT----- 98

```

```

Db 294 ALERKTDSPRYPRGSGNPKIALKIAEPQDSQCKIVSQEKELVQPFSSLPKVEYIA 353
QY 99 ----- 98
Db 354 RAGWTRDGKYAMAMFLDRPQOMLOLVLLPPALFIPSTENEBORLASARAVPRNVQPYVY 413
QY 99 ----- 98
Db 414 EEVTNWIMVNDIFYPPQSGEDBLCPRLANECKTGFCILYKTVANLKSQGYWSEPPS 473
QY 99 ----- 98
Db 474 PGEDEFKCPKEBIALTSGMEVILARHSGKIWNNEETKLVYFQGTQDTPLEHLLYVVSYE 533
QY 127 NPGEVTRLTDGYSGCCISQHCDFIISKYNQKNPHCVSLYKLSPEDDPTCKTEFWA 186
Db 534 AAGGIIVLTPRGFSHSCMSQNFDMFVSHSSYSTPCVHYVLTSGPDDDLPHKQRFWA 593
QY 187 TLIDSAGPLPDYTPPEIIFSPFSTGTFTLYGMLYKPHDLPQKKYPTVLFIYGPQVQLVN 246
Db 594 SMMEASCPDYPVPEIIFHFTSRDVLXGMLYKPHALQPKKHPVLFIYGGPQVQLVN 653
QY 247 NREKGVYFRLNTLASIGYVVVVIDNNGSCHRGKFEKGAFFYKMGQIEIDQVEGLQVLA 306
Db 654 NSFEGIKYRLNTLASIGYVVVVIDNNGSCHRGKFEKGAFFYKMGQIEIDQVEGLQVLA 713
QY 307 SRYDFIDLDVYGIGMSYGYLSLALMQRSDIFRYALAGAPYTLFIYDGYT 360
Db 714 EKXGFIIDSRVAIHGMSYGYLSLALMGLIHKQVFKVALAGAPYTMAYDGYT 767

RESULT 9
Q8BWT9 PRELIMINARY; PRT; 862 AA.
ID Q8BWT9
AC Q8BWT9;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DB Dipeptidyl peptidase 9 homolog.
GN DP9.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Liver;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK050021; BAC34034.1; -.
SQ SEQUENCE 862 AA; 98001 MW; B1D566E824A834B8 CRC64;

Query Match 62.3%; Score 1204.5; DB 11; Length 862;
Best Local Similarity 43.8%; Pred. No. 4.6e-101;
Matches 234; Conservative 53; Mismatches 72; Indels 175; Gaps 2;

QY 2 EDARSAGVATFVLOEFPDRYSGYWCPCAKETTPSGG-KILRIYEENDESVEIIVTSP 60
Db 233 DPKSAGVATFVLOEFPDRYSGYWCPCAKETTPSGG-KILRIYEENDESVEIIVTSP 292
QY 61 MLETRADSPRYPKTGTANPKVTFKMSIIMIDABGRIT----- 98
Db 293 ALERKTDSPRYPRGSGNPKIALKIAEPQDSQCKIVSQEKELVQPFSSLPKVEYIA 352
QY 99 ----- 98
Db 353 RAGWTRDGKYAMAMFLDRPQOMLOLVLLPPALFIPAVESEAQRQAAARAVKNAVQPYVY 412
QY 99 ----- 98

```

Db 413 EEVYNNWVNHDIFFHPPQAEAGODFCFLANCKTGFGHLYRVYELTKDYDWTBPLS 472

Qy 99 -----VDEVRLVYFEGTKDSPLHLLVYVY 126

Db 473 PTEDEFCPIKEEVALTSGEMEVLRSRHSKIMVNEQTKLVYFOGTQDTLEHLLVYVSYE 532

Qy 127 NPEVYRLTDRGVSHSCCISQHCDFISKYSNKNHCVSLYKLSPEDDPCKTKEFMA 186

Db 533 SAGEIVALTTLGFSHSCSMQSPFMYSHSVSTPEFCVAVNYKLSGDDDLKQPRFMA 592

Qy 187 TILDSAGBLPDYTPPELIFSESTGTGFTLYGMLYKPHDLQPKKYPTVLFYGGPOVOLVN 246

Db 593 SMMEANCPDYPPELIFHHTRADYQVLGYMICYKPHTLQGRKHPTVLFYGGPOVOLVN 652

Qy 247 NREKGVYFRNLNTLASIGYVYVINDRSGCHRLKEGAFKYMGOEIDDOVEGLQYLA 306

Db 653 NSFGIKYLRNLNTLASIGYVAVVINDRSGCORGLHFGALKNMGQVEIEDOVEGLQYLA 712

Qy 307 SRVYFDLDRVGIHGWMSGYGLSLMALMQRSDIFRVAIAGAPVTLMIFYDTGYT 360

Db 713 EKYGFDLSRVALIHGWMSGYGLSLMGLIHKQVFKVAILAGAPVTLMAVDTGYT 766

RESULT 10

Q8BVG4 PRELIMINARY; PRT; 862 AA.

ID Q8BVG4

AC Q8BVG4

DT 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Dipeptidyl peptidase 9 homolog.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CS7BL/6J; TISSUE=olfactory brain;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT Nature 420:563-573 (2002)."

RL Nature 420:563-573 (2002).

DR EMBL; AK078301; BAC37211.1; -

SQ SEQUENCE 862 AA; 98049 MW; B4FD3476B6F50030 CRC64;

Query Match 62.3%; Score 1204.5; DB 11; Length 862;

Best Local Similarity 43.8%; Pred. No. 4,6e-101;

Matches 234; Conservative 53; Mismatches 72; Indels 175; Gaps 2;

Qy 2 EDARSAGVATFVLOEEDRYSYWCPCAKETTPSG-KILRIYENDSEVYIHWISP 60

Db 233 DNPSAGVATFVLOEEDRYSYWCPCAKETTPSG-KILRIYENDSEVYIHWISP 292

Qy 61 MLETRRADSPRYPKTGTPANPKVTFKMEIMIDAEGRII----- 98

Db 293 ALBERKTDYRYRPTSGKNKIKLAELQTDHOKTVSSCEKLVOPRSSLPKRYEIA 352

Qy 99 ----- 98

Db 353 RAGWTRDGYAMMFLYRPOORLQVLLPPLFIPAVESBACROAARAVPKVQPFVY 412

Qy 99 ----- 98

Db 413 EEVYNNWVNHDIFFHPPQAEAGODFCFLANCKTGFGHLYRVYELTKDYDWTBPLS 472

Qy 99 -----VDEVRLVYFEGTKDSPLHLLVYVY 126

Db 473 PTEDEFCPIKEEVALTSGEMEVLRSRHSKIMVNEQTKLVYFOGTQDTLEHLLVYVSYE 532

Qy 127 NPEVYRLTDRGVSHSCCISQHCDFISKYSNKNHCVSLYKLSPEDDPCKTKEFMA 186

Db 533 SAGEIVALTTLGFSHSCSMQSPFMYSHSVSTPEFCVAVNYKLSGDDDLKQPRFMA 592

Qy 187 TILDSAGBLPDYTPPELIFSESTGTGFTLYGMLYKPHDLQPKKYPTVLFYGGPOVOLVN 246

Db 593 SMMEANCPDYPPELIFHHTRADYQVLGYMICYKPHTLQGRKHPTVLFYGGPOVOLVN 652

Qy 247 NREKGVYFRNLNTLASIGYVYVINDRSGCHRLKEGAFKYMGOEIDDOVEGLQYLA 306

Db 653 NSFGIKYLRNLNTLASIGYVAVVINDRSGCORGLHFGALKNMGQVEIEDOVEGLQYLA 712

Qy 307 SRVYFDLDRVGIHGWMSGYGLSLMALMQRSDIFRVAIAGAPVTLMIFYDTGYT 360

Db 713 EKYGFDLSRVALIHGWMSGYGLSLMGLIHKQVFKVAILAGAPVTLMAVDTGYT 766

RESULT 11

Q9BVM4 PRELIMINARY; PRT; 310 AA.

ID Q9BVM4

AC Q9BVM4

DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE Dipeptidyl peptidase 8 (Fragment).

GN DPP8.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=placenta;

RX MEDLINE=20467194; PubMed=11012666;

RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W., Gorrell M.D.;

RT "Cloning, expression and chromosomal localization of a novel human

RT dipeptidyl peptidase (DPP) IV homolog, DPP8."

RL Eur. J. Biochem. 267:6140-6150 (2000).

DR EMBL; AF221635; AAG29767.1; -

DR MEROPS; S09.018; -

DR InterPro; IPR001375; Peptidase_S9.

DR Pfam; PF00326; Peptidase_S9; 1.

FT NON TER 1 1

SQ SEQUENCE 310 AA; 35396 MW; BB87C34026D9C7AC CRC64;

Query Match 56.2%; Score 1086.5; DB 4; Length 310;

Best Local Similarity 80.6%; Pred. No. 6,7e-91;

Matches 204; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

Qy 108 FEGTKDSPLHLLVYVSVNPGVYRLTDRGVSHSCCISQHCDFISKYSNKNHCVSL 167

Db 1 FEGTKDSPLHLLVYVSVNPGVYRLTDRGVSHSCCISQHCDFISKYSNKNHCVSL 60

Qy 168 YKLSPEDDPCKTKEFWATILDSAGBLPDYTPPELIFSESTGTGFTLYGMLYKPHDLQPG 227

Db 61 YKLSPEDDPCKTKEFWATILDSAGBLPDYTPPELIFSESTGTGFTLYGMLYKPHDLQPG 120

Qy 228 KKYPVYLFYGGPOVOLVNNRKFVKYFRNLNTLASIGYVYVINDRSGCHRLKEGAFK 287

Db 121 KKYPVYLFYGGPO----- 134

Qy 288 YRMGOEIDDOVEGLQYLASRYFDLDRVGIHGWMSGYGLSLMALMQRSDIFRVAIAGA 347

Db 135 ---GOEIDDOVEGLQYLASRYFDLDRVGIHGWMSGYGLSLMALMQRSDIFRVAIAGA 191

Qy 348 PVTLMIFYDTGYT 360

Db 192 PVTLMIFYDTGYT 204

RESULT 12

Q9BVR3 PRELIMINARY; PRT; 439 AA.

ID Q9BVR3


```

QY 99 ----- 98
Db 121 HVPQSHBEIEFIASECKTGFRHLYKITSILKSKYKRSSGGLPAPSDKCPKEIEIA 180
QY 99 ----- 141
Db 181 ITSGEWEVLGRHGSNIQDEVARLVYFEGTKDSPLEHLYVSVVPCETVRLDRGSH 141
181 ITSGEWEVLGRHGSNIQDEVARLVYFEGTKDSPLEHLYVSVVPCETVRLDRGSH 240
QY 142 SCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTEFWATILDSAGPLDYTPP 201
241 SCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTEFWATILDSAGPLDYTPP 300
Db 202 EIFPSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVNNRPFKGVKFRLLNTLA 261
301 EIFPSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVNNRPFKGVKFRLLNTLA 340
QY 262 SLGYVVVVIDNRGSGHGLKEGAFKYMGGQIEIDQVEGLQYLASRYDFIDLDRVGIGHG 321
341 ----- 340
Db 341 ----- 340
QY 341 ----- 340
Db 341 ----- 340

```

RESULT 15

```

Q96NT8 . PRELIMINARY; PRT; 312 AA.
AC Q96NT8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein FLJ30094.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hiro Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Magatsuna M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054636; BAB70784.1; -.
DR InterPro; IPR001375; Peptidase_S9.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Hypothetical protein.
SQ
SEQUENCE 312 AA; 35518 MW; ABE940AFCS877717 CRC64;

```

Query Match 41.3%; Score 799; DB 4; Length 312;

Best Local Similarity 66.2%; Pred. No. 1.3e-64;

Matches 143; Conservative 31; Mismatches 42; Indels 0; Gaps 0;

```

QY 145 ISOHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTEFWATILDSAGPLDYTPP 204
Db 1 MSQNDPMFVSHSYSTPCVAVYKLSGDDPLHKKPRFWSMMEASCPDYPPEIF 60
QY 205 SFESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVNNRPFKGVKFRLLNTLA 264
Db 61 HHTRSQVLYGMLYKPHDLQPKKYPVLYFYGGPQVQVNNRPFKGVKFRLLNTLA 120
QY 265 YVVVVIDNRGSGHGLKEGAFKYMGGQIEIDQVEGLQYLASRYDFIDLDRVGIGHG 324
Db 121 YAVVVIDNRGSGHGLKEGAFKYMGGQIEIDQVEGLQYLAEKYGFTDLRVAIHGMSY 180
QY 325 GGYLSLMLMORSDIFRVAIAGAPVTLWIFDYTYT 360

```

```

Db 181 GGFSLMGLIHKEQVFKVAIACAPVWMAVDYTYT 216
Search completed: October 15, 2003, 17:13:47
Job time : 25.0243 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:15:46 ; Search time 3930.37 Seconds
(without alignments)
3747.092 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 BEDARAGVATFVLQEFPR.....RVAIAGAPVTLMIFYDTGYT 360

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=x1h
-O=/cgn2.1/US10.070464/runat 15102003 113553 24810/app query.fasta_1.2652
-DB=GenEmul -QPMF=fstep -SUFFIX=rge -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45
-DOCALLGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US10070464.@CGN 1 1 13561 @runat 15102003 113553 24810 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmul.*
1: gb_da.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_scs.*
28: em_un.*

29: em_vl.*
30: em_htg_hum.*
31: em_htg_in.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_gy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1833	100.0	1083	9	AF221637 Homo sapi
2	1880	97.3	4309	6	AX608737 Sequence
3	1836.5	95.0	2649	6	AX354795 Sequence
4	1836.5	95.0	2656	9	AY172659 Homo sapi
5	1836.5	95.0	2671	6	AX608725 Sequence
6	1836.5	95.0	3106	6	AX342633 Sequence
7	1836.5	95.0	3127	9	AF221634 Homo sapi
8	1836.5	95.0	3143	6	AX354793 Sequence
9	1836.5	95.0	4535	6	BC040203 Homo sapi
10	1816.5	94.0	4829	6	AX608735 Sequence
11	1756.5	90.9	4799	10	BC043124 Mus muscu
12	1534	79.4	3130	9	BC030688 Homo sapi
13	1528	79.0	2161	6	BD157001 Primer fo
14	1528	79.0	2161	6	AK027826 Homo sapi
15	1522	78.7	2510	6	AX338497 Sequence
16	1522	78.7	4685	6	AX608745 Sequence
17	1521.5	78.7	4676	6	AX608743 Sequence
18	1254.5	64.9	2668	6	AX405771 Sequence
19	1254.5	64.9	2842	6	AX405770 Sequence
20	1230	63.6	2778	9	AK000290 Homo sapi
21	1229	63.6	4523	6	AX608731 Sequence
22	1220.5	63.1	2602	9	AX608727 Sequence
23	1220.5	63.1	2617	6	AX608727 Sequence
24	1220.5	63.1	2676	6	AX524928 Sequence
25	1220.5	63.1	3000	9	AF542510 Homo sapi
26	1220.5	63.1	3243	9	HSMB05448 Homo sapi
27	1220.5	63.1	3716	6	AX480934 Sequence
28	1220.5	63.1	4219	6	AX608751 Sequence
29	1220.5	63.1	4232	9	BC037948 Homo sapi
30	1220.5	63.1	4295	9	AF452102 Homo sapi
31	1220.5	63.1	4302	6	AX608747 Sequence
32	1201	62.1	4180	6	AX608759 Sequence
33	1201	62.1	4263	6	AX608757 Sequence
34	1132.5	58.6	4076	6	AX608755 Sequence
35	1132.5	58.6	4159	6	AX608753 Sequence
36	1113	57.6	4037	6	AX608763 Sequence
37	1113	57.6	4120	6	AX608761 Sequence
38	1089	56.3	753	6	AX524935 Sequence
39	1088.5	56.2	1197	9	AF221635 Homo sapi
40	984	50.9	2546	9	BC000970 Homo sapi
41	975	50.4	2261	6	AX713363 Sequence
42	975	50.4	2261	6	AK054656 Homo sapi
43	860.5	44.0	1669	9	AF221636 Homo sapi
44	735	38.0	1080	6	AX524932 Sequence
45	718.5	37.2	3740	3	BT001499 Drosophila

ALIGNMENTS

RESULT 1

AF221637 1083 bp mRNA linear PRI 05-NOV-2000
 LOCUS Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, partial cds,
 DEFINITION alternatively spliced.
 ACCESSION AF221637
 VERSION AF221637.1 GI:11095193
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 1083)
 AUTHORS Abbott, C.A., Yu, D., McCaughan, G.W., and Gorrell, M.D.
 TITLE Cloning, expression and chromosomal localization of a novel human dipeptidyl peptidase (DPP) IV homolog, DPP8
 JOURNAL Eur J Biochem. 267 (20), 6140-6150 (2000)
 MEDLINE 20467194
 PUBMED 11012666
 REFERENCE 2 (bases 1 to 1083)
 AUTHORS Abbott, C.A., Yu, D., McCaughan, G.W. and Gorrell, M.D.
 TITLE Direct Submission
 JOURNAL Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
 Centenary Institute of Cell Biology and Cancer Medicine,
 Locked Bag No. 6, Newtown, Sydney, NSW 2042, Australia
 FEATURES
 source
 1. .1083
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q22"
 /clone="scPBCd3-3-1"
 /cell_type="PHA stimulated PBMC"
 <1. >1083
 /gene="DPP8"
 <1. >1083
 /gene="DPP8"
 /note="serine protease; member of prolyl oligopeptidase S9b family belonging to enzyme clan SC; additional splice variants of this mRNA are found in GenBank Accession Numbers AF221634, AF221635 and AF221636"
 /codon_start=2
 /product="dipeptidyl peptidase 8"
 /protein_id="P46297.9.1"
 /db_xref="GI:11095194"
 /translacion="BEDARAGVATFVLOEFPDYSGYWMCPRATTPSGKILRIILY
 EENDESFEIIVHTSPMLERADSFERYPTGTANPKVTFKMSBIMIDAGRIIVBEV
 RLIVFEGTDSPLBHLVYVSVNPGEVRLIDRGVSHSCISOHDFISKYSNK
 NPKVDSLYKLSPEPDPTCKTKERPMATILDSAGLIDPVPPEIISPSFTGLTYGKL
 YKPHDLPGRKRTIVLFITGGPOVLVNNFKGVKFRRLTSLGIVYVYIDNRSC
 HRGLKEGAFKRYKQGLEIDQVQGLVLSRYPDIDLDRVGIHWSYGYSLSLMLM
 QRSDFEVAIAGAPVTLIMIFYDTGT"
 BASE COUNT 312 a 213 c 247 g 311 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.24e-191 Length: 1083
 Score: 1933.00 Matches: 360
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-10-070-464-7 (1-360) x AF221637 (1-1083)
 QY 1 GIUGIUSPALAAGSSEALAGLYVALAATHRPHVEALLENGINGIUGIUPHEASPARQ 20
 DB 2 GAAGAAGATGCCAATACGCTGAGAGCTGCTTGTCTCCAGAAAGATTTCATAGA 61
 QY 21 TYSERGLYTYRTPTTPCYSPFOLYSAIAGIUTHRTPROSERGLYGLYISILEU 40
 DB 62 TATCTGGCTATTGTGTGTCCAAAGCTGAACAACCTCCAGTGTGTGTAATAATCTT 121

QY 41 ArgIleuTYRGIUGIUSASPSGIUSERGIUVALGIUILEIIEHISVALTHRSERPRO 60
 DB 122 AGAATTCATATGAAGAAATGATCAATCTGAGGCGGAATATTCATGTTACATCCCT 181
 QY 63 MetIeGIUTHRARARALASPSERPHEARGYPROLYTHRGYTHRALASNPRO 80
 DB 182 ATGTTGGAAACAGAGGCGAGATTCATTCGCTTATCCTAAGACAGGTACACCAATCCT 241
 QY 81 LYSVALTHRPHLYEMESERGIULEMETILEASPAIAGIUGIYARGILEIIVLASP 100
 DB 242 AAAGTCACCTTTAAATGTCAGAAATAATGATTGATGCTGAAGAAAGATCAATAGTAT 301
 QY 101 GIUVALARGARLEUVALTYRPHGIUGIYTHLYASPSERPROLENGIUNISHISLEU 120
 DB 302 GAAGTCAGAGGCTGGATATTTTGAGGACCAAGACTCCCTTTAAGCATACCTG 361
 QY 121 TYRVALIASERTYRVALASNPROGIUGIUALTHYARGLEUTHRASPARGIYYSER 140
 DB 362 TACGTATGTCAGTACGTAAATCCTGGAGAGTGACAAAGCTGACCGTGGCTACTCA 421
 QY 141 HISSECYSPYSESERGINHISCYASAPHPHEIIESERTYSTYRSERASNGIUNLYS 160
 DB 422 CATTCCTGCTGCATCGACGACGCTGACTCTTTATAGTAAGTATGATTAACCGAAG 481
 QY 161 ASNPROIHCYVALSERLEUTHYRILYSLEUSERSERPROGIUSASPSERPROTHRCYSLYS 180
 DB 482 AATCCACACTGTGTGTCCTTTACAGACTATCAAGCTCGAAGATGACCAACTTGGAA 541
 QY 181 THLYSGIUPHETRPALATHRIILEUASPSERALGIYPROLEUPROASPTYRTHRPRO 200
 DB 542 ACNAAAGAAATTTGGGCCACCAATTTGGATTGACGAGGTCCTCTCTGACTATACCT 601
 QY 201 PROGIULIPESESERHEGLUSERTHRTYRTHYRPHETHLEUTHYRGLYMETLEUTHYRYS 220
 DB 602 CCAGAAATTTCTCTTTTAAAGTACTACTGATGATTAAGTATGAGGAGCTCTACAG 661
 QY 221 PROHISASPLEUGINPROGIULYSYSTYRPROTHRVALLEUPHEILEYRGYIYPRO 240
 DB 662 CCTCATGATCTACAGCTCGAAGAAATATCCATGCTGTGTCATATATGAGGCTCT 721
 QY 241 GINVALGILUEVALASNSANARGPHELYGIVALLYSTYRPHEARGLEUASNTHLEU 260
 DB 722 CAGGTGCAAGTGTGATATATCGTTTAAAGGAGTCAAGATTTCCGCTTGAATACCTCA 781
 QY 261 ALISERLEUGIYTYRVALIVALIILEASPSARGLYSERCYSHISARGGLYLEU 280
 DB 782 GCCTCTCTAGTATATGTGTGATGATGACAAACAGGGGATCTGTCACCGAGGCTCT 841
 QY 281 LYSPEGLUGIVALAPHELYSTYRILYSMEGLGINILEGIULIIEASPSARGINVALGIU 300
 DB 842 AATTTGAAAGCCCTTTTAAATATTAATAATGCGTCAATAATTAATGATCGATCGGTGAA 901
 QY 301 GILYLEUGINTYRLEUALASERAGTYRASPPHEILEASPLEUSPARGLYIILEHIS 320
 DB 902 GAACCTCAATATATCTAGCTTCTCGATATGATTAATGATCACTTATGCGGCATCTCAC 961
 QY 321 GLYTPSERTYRGLYTYRILEUSERLEUETILEUWETGINARSEARPIIPEHE 340
 DB 962 GGCCTGCTCATAGAGATACCTCTCCCTATGACATTAATGACAGAGGTAGATATCTTC 1021
 QY 341 ARGVALAIATIEALAGIYALAPROVALTHRLEUTRPIIPEHETRYASPTHRGIYTYRTHR 360
 DB 1022 AGGTTCTATTTGCTGGGCGCCAGTCACTCTGGATCTTCTATGATACAGATACAGC 1081
 RESULT 2
 AX608737 4309 bp DNA linear PAT 17-FEB-2003
 LOCUS Sequence 14 from Patent WO02311134.
 DEFINITION AX608737
 ACCESSION AX608737
 VERSION AX608737.1 GI:28404304
 KEYWORDS Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
AUTHORS Qi, S., Akinsanya, K.O., Riviere, P.J. and Junien, J.L.
TITLE Novel serine protease genes related to dppiv
JOURNAL Patent: WO 0231134-A 14 18-APR-2002;
Ferring BV (NL)

FEATURES
source Location/Qualifiers
1..4309
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 1304 a 790 c 907 g 1308 t
ORIGIN

Alignment Scores:
Pred. No.: 7,46e-185 Length: 4309
Score: 1880.00 Matches: 358
Percent Similarity: 98.90% Conservative: 0
Best Local Similarity: 98.90% Mismatches: 2
Query Match: 97.26% Indels: 3
DB: Gaps: 0

US-10-070-464-7 (1-360) x AX608737 (1-4309)

QY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluIleuPheAspArg 20
DB 943 GAAGAAGATGCCAGATCAGCTGAGTCGCTACCTTTGTTCTCCAGAAAGATTGATTAGA 1002

QY 21 TyrSerGlyTyrTrpTyrCysProlysalGluThrThrProSerGlyGlyLysIleu 40
DB 1003 TATTCGGCTATTGGTGCTGCTCAAAAGCTGAACAACATCCAGTGGTGTAAATTCCT 1062

QY 41 ArgGluLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
DB 1063 ACAAATCTATATGAAGAAATAATGATGAGGTGAATAATTCATGTTACATCCCT 1122

QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
DB 1123 ATGTTGGAAACAAGAGGCGCAGATTCATTCCTGTAACCTTAACCAAGGTACAGGAATCCT 1182

QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
DB 1183 AAAGTCACCTTTAAAGATGTCAGAAATAATGATGATGCGAAGAGATCCAA-GTTGAT 1241

QY 101 GluValArgArgLeuValTyrPheGluGluThrLysAspSerProLeuGluHisIleu 120
DB 1242 GAAGTCAGAGGCTGATATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACTG 1301

QY 121 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 140
DB 1302 TACGTAGTCAGTTACCTAAATCTCTGAGAGGTACCAAGGTGATCACTGCTGCTACTCA 1361

QY 141 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 160
DB 1362 CATTTCTGCTGATCAGTCAGCACTGTGACTTCTTATTAAGTAAAGTATGTAACCAAG 1421

QY 161 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 180
DB 1422 AATCCACACTGTGTGCTTCTTACAGCTATCAAGTCCGAAATGACCAACTTCGAAA 1481

QY 181 ThrLysGluPheTyrPheAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 200
DB 1482 ACAAGAGAAATTTGGGCCACCAATTTGATTCAGAGGTCCTCTCTGACTATACCTCT 1541

QY 201 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
DB 1542 CCAGAAATTTCTCTTTTGAAGTACTACGATTTACATTTGATGAGATGCTCTCAAG 1601

QY 221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly-P 240
DB 1602 CCTCATGATCTACAGCTGGAAGAAATATCTTACTGTGCTGTTCAATATATGATGATCTC 1661

QY 240 rGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrL 260
DB 1662 CTCAGGTGACAGTGGTGAATATCGGTTTAAAGGAGTCAAGATATTCGCTTGAATACCC 1721

QY 260 eulAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCysHisArgGlyL 280
DB 1722 TAGCTCTCTAGGTTATGTGTGTAGTATGAGCAAGAGGATCTGTCCAGAGGAC 1781

QY 280 eulYpPheGluGlyValAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValG 300
DB 1782 TTTAAATTTGAAGGCCCTTTAAATTTAAATGGCTCAATTAATGATGATCAGGTGG 1841

QY 300 lueGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyTLeH 320
DB 1842 AAGACTCCAAATATCTAGCTTCGATATGATTTGATTAATGATCTTATGATACAGATACA 1901

QY 320 lseGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIleP 340
DB 1902 ACGGCTGGCTCATGAGAGATACCTCTCTGATGCAATTAATGACAGAGTCAGATATCT 1961

QY 340 heArgValAlaIleAlaGlyValAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrT 360
DB 1962 TCAGGATTCATTTGCTGGGCCCCAGTCACCTCTGTGATCTTCTATGATACAGATACA 2021

QY 360 hr 360
DB 2022 CG 2023

RESULT 3
AX354795
LOCUS AX354795 2649 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 3 from Patent WO0179473.
ACCESSION AX354795
VERSION AX354795.1 GI:18619528
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
AUTHORS Meyers, R.A. and Williamson, M.
TITLE 21953, a human prolyl oligopeptidase family member and uses thereof
JOURNAL Patent: WO 0179473-A 3 25-OCT-2001;
Millennium Pharmaceuticals, Inc. (US)

FEATURES
source Location/Qualifiers
1..2649
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 803 a 514 c 586 g 746 t
ORIGIN

Alignment Scores:
Pred. No.: 1,36e-180 Length: 2649
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: Gaps: 1

US-10-070-464-7 (1-360) x AX354795 (1-2649)

QY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluIleuPheAspArg 20
DB 730 GAAGAAGATGCCAGATCAGCTGAGTCGCTACCTTTGTTCTCCAGAAAGATTGATTAGA 789

QY 21 TyrSerGlyTyrTrpTyrCysProlysalGluThrThrProSerGlyGlyLysIleu 40
DB 790 TATTCGGCTATTGGTGCTGCTCAAAAGCTGAACAACATCCAGTGGTGTAAATTCCT 849

QY 41 ArgGluLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
DB 1602 CCTCATGATCTACAGCTGGAAGAAATATCTTACTGTGCTGTTCAATATATGATGATCTC 1661

Db AGATTCTATATGAGAAATGATGATCTGAGGTGAGAAATTATTCATGTTACATCCCT 909
 Qy 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProlyThryAlaAsnPro 80
 Db 910 ATGTGGAAACAGAGGGGAGATTCATCCGTTATCCCTAAACAGAGTACAGCAATCCCT 969
 Qy 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
 Db 970 AAGTCTACTTTTAAGATGTCAGAAATTAAGATTGATGCTGAAAGAAAGATCATAGATGTC 1029
 Qy 98 ----- 98
 Db 1030 ATAGATTAAGAACTAATTCACCTTTTGAGATTCTATTGTAAGAGATTGAATATATGTC 1089
 Qy 98 ----- 98
 Db 1090 AGAGCTGATGATCTCCCTGAGGAAATATGCTTGTCATCTACTAGATGCTCCAG 1149
 Qy 98 ----- 98
 Db 1150 ACTCGCTGACATAGTGTGATCTCACTGAAATTATTAATCCAGTAGAAGATGATGTT 1209
 Qy 98 ----- 98
 Db 1210 ATGAAAGGAGAGACTCAATGAGTCAGTGCCTGATTCGTGACGCCACTAATTATCTAT 1269
 Qy 98 ----- 98
 Db 1270 GAAGAAACACAGACATCTGGATAAATATCCATGACATCTTCATGTTTCCCAAGT 1329
 Qy 98 ----- 98
 Db 1330 CACGAAGAGAAATTAGATTATTTTGGCTCTGAATGCAAAACAGTTTCCCTCATTTA 1389
 Qy 98 ----- 98
 Db 1390 TACAAATTATCATCTATTTTAAAGAAAGCAATATAACGATCCAGTGTGGCTGCT 1449
 Qy 98 ----- 98
 Db 1450 GCTCCAGTGAATTCAGTGTCTCTATCAAGAGAGATAGCAATTAACAGTGTGAATG 1509
 Qy 99 ----- ValAspGluValArgArgLeuValTyr 107
 Db 1510 GAAGTCTTGCCGCGATGATCTAATATCCAGTTGATGAATCAGGAAGCGGTATAT 1569
 Qy 108 PheGluGlyThrLysAspSerProLeuGluHisLysLeuTyrValValAsn 127
 Db 1570 TTGGAAGCACCAAGACTCCCTTTAGAGCATCACCTGTACGTACGTTACGTAAAT 1629
 Qy 128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysLysSerGln 147
 Db 1630 CCTGGAGAGGTGACAGGCTGACTGACCGGTGCTACTCACATTTCTGCTGATCAGT 1689
 Qy 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
 Db 1690 CATGTGACTTCTTTTAAGTAAGTATAGTAACACAGAGATCCACATGTGTGCTCT 1749
 Qy 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPalatnr 187
 Db 1750 TACAGACTATCAAGTCTCTGAAGATGACCCCACTGCAAAACAAAGAAATTTTGGGCACC 1809
 Qy 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
 Db 1810 ATTTTGATTCACAGGCTCTCTCTCTGACTATATCTCTCCAGAAATTTCTCTTTTGA 1869
 Qy 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 Db 1870 AGTACTACTGATTTTACATTGATGATGGATCTCTACAGAGCTCATGATCTACAGCTGGA 1929
 Qy 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnAlaGlnLeuValAsn 247
 Db 1930 AAGAAATATCTACTGTGCTGTCTATATATGATGATGCTCAGGTGCGAGTTGGAAATAT 1989

Qy 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 Db 1990 CGTTTAAAGGAGTCAAGATATTCGCTTGAAATACCTTAGCTCTTAGGTTATGTGTT 2049
 Qy 268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyValaPheLys 287
 Db 2050 GTAGTATGACAAACAGGGATCTCTGACCGAGGCTTAATTTGAAAGCCCTTTAA 2109
 Qy 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
 Db 2110 TATTAATATGGCTCAATAGAAATTAACGATCAGGTGAGAGACTCCAAATATCTAGCTTCT 2169
 Qy 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyThrSerTyrGlyTyr 327
 Db 2170 CGATATGATTTATTAATGATGATGTCGTGGCATCAGCGCTGCTCATGAGATAC 2229
 Qy 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValaIleAlaGlyAla 347
 Db 2230 CTCTCCCTGATGGCATTAATGACAGAGTCAAGATCTTCAGGGTTCATGCTGAGGCC 2289
 Qy 348 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
 Db 2290 CCAGTCACCTCTGTGATCTTATGATATACAGATACACG 2328
 RESULT 4
 LOCUS AY172659 2656 bp mRNA linear PRI 08-JAN-2003
 DEFINITION Homo sapiens dipeptidyl peptidase IV-related protein-1 (DPP1)
 ACCESSION AY172659
 VERSION AY172659
 KEYWORDS mRNA, complete cds.
 SOURCE AY172659.1 GI:27549549
 ORGANISM Homo sapiens (human)
 REFERENCE 1 (bases 1 to 2656)
 AUTHORS Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
 TITLE Novel Serine Protease Genes Related To DPP1V
 JOURNAL Patent: US (WO 0231134) -A 18-APR-2002;
 AUTHORS Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
 TITLE Direct Submission
 JOURNAL Submitted (04-NOV-2002) Ferring Research Institute, 3550 General
 Atomic Ct., San Diego, CA 92121, USA
 FEATURES
 source
 1..2656
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q22.1-q22.2"
 1..2656
 /gene="DPP1"
 8..2656
 /gene="DPP1"
 /codon_start=1
 /product="dipeptidyl peptidase IV-related protein-1"
 /protein_id="AA017261.1"
 /db_xref="GI:27549550"
 /translation="MAAMETQLGVEIPETADCEENIESDRPKLEPFYVERYSWQ
 LKKLADTRKGYWMAKAPHPFMFKRNDPPGSHDRITVYAMSGENKNTFYSEI
 PKTIRAAVIMSKPRLIDLRQATLDYGGYSEELIRKRTIGYAGISYVHOSSG
 TPLPQSGITVYKKGPGFTQQLPRNVLTSCNIMDKLCPADPDWIAFIHSN
 DIMSIVTRERRLTYVNELANMEEDARSAGATFVQLQEFDRYSGYWCPEKAIT
 PGGKILRLIYENDESEVEIITHVSPMLERADSFYKGTGNPKYTFKMSLMI
 DABGRIDIVDKELIOPFEILFEGVEYIARAQMTPEGKYAMGILDRSQRLOVLS
 PELFIPVEDDVMOROLIESVDSYPLIIEYETDINWINDIPHVPOSHSEIEF
 ITASECKGRFRLYKITSILSKSKYRSGGGLPDSDFCPILKEEIAIRSGWEIVG
 HSNTOUVERLIVFEGTKDSPLKHHLYVYVYVNPGEVTRLIDRGYSKCCISOHD
 FRTSKSNQKPHCYSLKUSPEDDPTKTEFRNTILDSKGLDLDIYPPPIFESES
 LTGFLYGLMYKPHDLQPEKPIPTVLFYGGPQVQLVNNRFGVKYFRINTLASLGV

BASE COUNT 804 a 517 c 588 g 747 t
ORIGIN

VVIDNRGSGHGLKEGAFKYNKGIEIDDOVBGLQYLAISYDFLIDREVGJHGSY
GGYISLIMAMORSDIFRVAIAGAPVTIMFYDGYTERYMGHPDONOGYILSVAMQ
AEKPESEPNLLHLHGFLENDENHFAHRSILSFLVRAGKPYDQIYQERHSRIVES
GEHYELHLHYLDENLGSRIALAKVI"

Alignment Scores:

Pred. No.: 1,36e-180 Length: 2656
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: Gaps: 1

US-10-070-464-7 (1-360) x AY172659 (1-2656)

```
QY      1  GUGUUSPALAATGserAaglyValAlaThrPheValLeuGlnGluGluPheAspArg 20
DB      737  GAAGAAATGCGCAATCAGCTGAGTGGCTACCTTGTCTCCAAAGAAATTTGATAGA 796
QY      21  TySerGlyTyrTrpTPCySProLySAlaGluThrThrProSerGlyGlyLeu 40
DB      797  TATTCTGCTATTGAGTGTGTCGAAAGCTGAAACACTCCAGTGGTGAATTTCTT 856
QY      41  ArgIleuTyrgIuGluuAsnAspGluSerGluValGluIleHisValThrSerPro 60
DB      857  AGAATTCCTAATGAAAGAAATGATGAAATCTGAGGTGGAATTAATTCATGTTACCTCCCT 916
QY      61  MetLeuGluThrArgAlaAspSerPheArgTyrProLySThrGlyThrAlaAsnPro 80
DB      917  ATGTTGAAACCAAGAGGCGAGATTCAATCCGTTATCCTAAACAGGTACGCAAAATCCT 976
QY      81  LysValAlaThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
DB      977  AAATGCTCTTTTAAGATGTCAGAAATATATGATTAATGCTGAAGAGATCATAGATGTC 1036
QY      98  ----- 98
DB      1037  ATAGATAAGAACTAAATTCAACTTTTGAGATTCATATTGAAGAGTTGATATATTGCC 1096
QY      98  ----- 98
DB      1097  AGAGCTGATGAGTACTCCTGAGGAAATATGCTTGCTTCATCTAGATCGCTCCAG 1156
QY      98  ----- 98
DB      1157  ACTGCGCTGCAGATAGTGTGATCTCACTGAATTATTATCCAGTAGAAGATGATGTT 1216
QY      98  ----- 98
DB      1217  ATGGAAGAGCAGAGTCAATGAGTCAGTGCCTGATTCGTGAGCGCACATAATTATCTAT 1276
QY      98  ----- 98
DB      1277  GAAGAAACACACAGCATCTGATTAATATCCATGACATCTTTCAATGTTTTTCCCAAGT 1336
QY      98  ----- 98
DB      1337  CACGAGAGAGAAATTGAGTTATTTTGGCTCTGAATGCAAAACAGTTTCCGTCATTTA 1396
QY      98  ----- 98
DB      1397  TACAAATAATCATCTATTTTAAAGAAAGCAAAATATAAACGATCCAGTGTGGCTGCT 1456
QY      98  ----- 98
DB      1457  GCTCCAAAGTATTTCAAGTGTCTTATCAAGAGAGATAGCAATTTACAGTGTGAATGG 1516
QY      99  -----ValaspGluValArgArgLeuValTyr 107
DB      1517  GAAATGTTGGCGGCGCATGATCTAATATCCAAAGTTATGAAGTCAGAAAGCTGTATAT 1576
```

```
QY      108  PheGluGlyThrLysAspSerProLeuGlnHisIleuTyrValIleSerTyrValAsn 127
DB      1507  TTGAGAGGCAACCAAGACTCCCTTTAGACATCACCTGACAGTACAGTTACGTAAAT 1636
QY      128  ProGlyValValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
DB      1637  CTGAGAGAGGTGACCAAGGCTGACTGACCGGTGGCTACTCAATCTTCTGATAGTCAG 1696
QY      148  HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
DB      1697  CACTGTACTCTTCTTTAATAGTAAGTATAGTACCAAGAAATCCACACTGTGTCTCCTT 1756
QY      168  TyrIleuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 187
DB      1757  TACAAAGCTATCAAGTCTCTGAAGATGACCCCAACTGCAAAACAAAGAAATTTGGGCCAC 1816
QY      188  IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
DB      1817  AYTATTGATTCAGAGGTCCTCTTCTGACTATACCTCCAGAAATTTTCTTTTGAA 1876
QY      208  SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
DB      1877  AGTACTACTGATTTTACATGTATGATGATGCTCTACAAAGCTCATGATCTACAGCTCGA 1936
QY      228  LysLysTyrProThrValIleuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 247
DB      1937  AAGAAATATCTCACTGCTGCTGCTCATATATGCTGCTCAGGTGCACTGATGATAT 1996
QY      248  ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuIleSerLeuGlyTyrValVal 267
DB      1997  CGATTTAAGAGCAAGATATTCGCTTGAAATCCCTAGGCTCTCTAGGTATAGTGT 2056
QY      268  ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluValAlaPheLys 287
DB      2057  GTAGTATAGCAACCAAGGATCTCTGACCGAGGCTTAATTTGAAGGCGCTTTAA 2116
QY      288  TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
DB      2117  TATATAATGGGTCAAAATAGAAATTCAGATCAGGTGAAAGGACTCCATATCTGCTTCT 2176
QY      308  ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyTyr 327
DB      2177  CGATATATTTTCATGCTTAGATCGTGTGGGATCCACGCGTGTCTTATGAGAGATAC 2236
QY      328  LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
DB      2237  CTCCTCCATGATGATTAATGACAGAGTCAGATATCTTCAGGGTTGCTATTTGCTGGGCC 2296
QY      348  ProValThrLeuTyrPheTyrAspThrGlyTyrThr 360
DB      2297  CCACTCACTCTGTGATCTTCTATGATACAGGATACAG 2335
```

RESULT 5

AX608725

LOCUS

DEFINITION

AX608725

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .2671
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 805 a 524 c 594 g 748 t
ORIGIN

Alignment Scores:

Pred. No.: 1,37e-180 Length: 2671
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
Gaps: 1

US-10-070-464-7 (1-360) X AX608725 (1-2671)

```

QY      1  GUGUASpalaagserlaaglyvalalathrphelalleuglingleupheasparg 20
DB      737  GAAGAAATGCCAATCAGCTGAGTGCCTACCTTGTCTCCAAAGAAATTTGATAGA 796
QY      21  Tysersglytrtpcyeprolyvalagluhrthrproserglylyleu 40
DB      797  TATTCGCTATTGCTGCTGCCAAAGCTGAAACACTCCAGTGGTGAATAATCTT 856
QY      41  Argilleuhyrglingleuasnpsglusergluvalglullehsvalthrserpro 60
DB      857  AGAATTCATATGAAGAAATGATGAATCTGAGGTGAATTTATTCATGTTACATCCCT 916
QY      61  Metleugluhrargalaaspsrpheargtyrprolysthrnglythrthalaasnpro 80
DB      917  ATGTTGAAACAAAGAGGAGGAGATTCATCCGTTATCTTAACAGGTACGCAAAATCT 976
QY      81  Lysvalthrphelysmetserglulleuethleaspalagluhyrargille- 98
DB      977  AAATCTACTTTAAGATGTCGAATAATATGATTGCTGAAGAGGATCATAGATGTC 1036
QY      98  ----- 98
DB      1037  ATAGATAAGAACTAATCAACTTTTGAGATTCATTTGAAGAGTTGAATATATGCC 1096
QY      98  ----- 98
DB      1097  AGAGCTGATGGACTCCTGAGGGAATAATGCTTGCTTCATCTACTAGATGCTCCAG 1156
QY      98  ----- 98
DB      1157  ACTGCGCTGAGATAGTGTGATCTCACCTGAATTTATCCAGTAGAAGATGATGT 1216
QY      98  ----- 98
DB      1217  ATGAAAGGACAGAGACTGATTGAGTCACTGCTGATTCTGTGACGCCACTAATTATCTAT 1276
QY      98  ----- 98
DB      1277  GAAGAAACACAGACATCTGATTAATATCCATGACATCTTTTCCCAAGT 1336
QY      98  ----- 98
DB      1337  CACGAAGAGAAATGAGTTATTTTCTCTGATGCAAAACAGTTCCGTCATTTTA 1396
QY      98  ----- 98
DB      1397  TACAAATTTACATCTATTTTAAAGAAACAAATATTAACGATCCAGTGGGCTGCT 1456
QY      98  ----- 98
DB      1457  GCTCCAAAGTATTCAGATGTCCTATCAAGAGAGATGAGCAATTACAGTGGGAATGG 1516
QY      99  -----Valaspgluvalargyleuvaltyr 107
DB      1517  GAAGTCTTGGCCGCGATGATCTAATATCCAAAGTGTAGAGTCAAGGCTGGTATAT 1576
QY      108  Phegluglythrlyaspserproleugluhrshisluhyrvalaserlyrvalasn 127
DB      1577  TTGAAGGACACCAAGACTCCCTTTAGAGCATCTGCTGACGTACGTACGTTACGTAAAT 1636

```

```

QY      128  ProgluvalthrargleuThrapparglytyrserhissercyscysileserghn 147
DB      1637  CCTGGAGAGGTGACCAAGGCTGACTGACCGTGGCTACTCATTCTTCTGCATCAGCAG 1696
QY      148  HisCysaspPhepheiLeserlyrSerasnGlnLysasnProhisCysvalSerleu 167
DB      1697  CACTGTGACTCTTTTAAAGTAAAGTATAGTAAACAGAAAGATCCAGCTGTGTCTCTT 1756
QY      168  TyrlyleuSerSerProgluaspaspProthrcyslysthrlysgluhrthralathr 187
DB      1757  TACAGGCTATCAAGTCTCGAAGATGAGCCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1816
QY      188  IleuaspserlaaglyrproleuproaspdyrthrproprogluilepheSerPheglu 207
DB      1817  ATTTGGATTCAGAGGCTCTCTCTGACTATACCTCCAGAAATTTTCTTTTGA 1876
QY      208  SerThrThrghlypethrleuhyrlymetleuhyrlyspProhisaspLeuGlnProgly 227
DB      1877  AGTACTATCGAATTTACATTTGATAGGATGCTGTACAGGCTCATGATCTACAGCTGGA 1936
QY      228  LyslystyrProthrvalleupheiletyrGlyGlyProGlnValGlnleuValAsn 247
DB      1937  AAGAAATATCTCTGCTGCTGCTCATATATGTGTGCTCAGAGTGTGTGAATAT 1996
QY      248  ArgPhehyrGlyVallystyrPheargLeuasnThrleuAlaserLeuGlyTyrrVal 267
DB      1997  CGATTTAAAGAGCAAGTATTTCCGCTTGAATACCTTAGCTCTCTAGTTATGTGTT 2056
QY      268  ValvalilleaspasnarglyserCysHisarglyleuLysPhegluValaPheLys 287
DB      2057  GTAGTATAGCAACACAGGGGATCCTGTACCGAGGGCTTAATTTGAAGGCGCCTTTAA 2116
QY      288  TyrlsmetGlyGlnleGluilleaspasglvalGluGlyLeuGlnTyrrleuAlaser 307
DB      2117  TATTAATGGGTCAATTAAGAAATTGAAGATCAGGTGAAGAGCACTCCAAATCTAGCTCT 2176
QY      308  ArgTyaspPheilleaspLeuaspargValGlyIleHisGlyTPSerTyrrGlyTyrr 327
DB      2177  CGATATATTTTCACTTATGATCGTGTGGGATCCAGGCTGTCTTATGAGAGATAC 2236
QY      328  LeuSerleuMetAlaMetGlnArgSeraspIlePheargValAlaIleAlaGlyAla 347
DB      2237  CTCTCCCTGATGGCATTAATGACAGAGGTGATATCTTACAGGTTGTGCTATGTGGGGCC 2296
QY      348  ProvalthrleuThrIlePheTyrrAspThrGlyTyrrthr 360
DB      2297  CCAGTCACTGTGTGATCTTCTATGATACAGATACAG 2335

```

RESULT 6
AX342633 3106 bp DNA linear PAT 12-JAN-2002
LOCUS
DEFINITION Sequence 30 from Patent WO0198468.
ACCESSION AX342633
VERSION AX342633.1 GI:18152030
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
Yue,H., Elliot,V.S., Gandhi,A.R., Lal,P., Au-Young,J.,
Tribouley,C.M., Deleogene,A.M., Baughn,M.R., Nguyen,D.B., Lee,E.A.,
Hafalla,A., Khan,F.A., Wallia,N.K., Yao,M.G., Lu,D.A., Patterson,C.,
Tang,Y.T., Walsh,R.T., Azimzai,Y., Ramkumar,J., Xu,Y. and Reddy,R.
Patent: WO 0198468-A 30 27-DEC-2001;
Incyte Genomics, Inc. (US)
Location/Qualifiers
1..3106
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 7160544CB1"
BASE COUNT 928 a 633 c 704 g 841 t

ORIGIN

Alignment Scores:

Pred. No.: 1,65e-180 Length: 3106
 Score: 1836.50 Matches: 360
 Percent Similarity: 67.548 Conservative: 0
 Best Local Similarity: 67.548 Mismatches: 0
 Query Match: 95.014 Indels: 173
 DB: 6 Gaps: 1

US-10-070-464-7 (1-360) x AX342633 (1-3106)

QY 1 GIUGIAspAlaArgSerIAGIyValAlaThrPheValLeuGlnGluPheAspArg 20
 DB 932 GAAGAAGATGCCAGATCAGCTGAGTGGCTACCTTGTCTCCAAAGAAATTTGATTACA 991
 QY 21 TYSerGIyTyrTyrPyrCysProLysAlaGluThrThrProSerGIyGlyLysIleLeu 40
 DB 992 TATTCGGCTATTGGTGTGTCTCCAAAGCTGAACACTCCAGTGTGTAAATTCCT 1051
 QY 41 ArgIleLeuTyrGlnGluAsnAspGluSerGluValGlnIleIleHisValThrSerPro 60
 DB 1052 AGAATTCATATGAAGAATAATGATGATCTGAGGTGAATATTATCATGTATCATCCCT 1111
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
 DB 1112 ATGTGGAAACAAGAGGGGAGATTCATTCCTTATCTTAAACAGGTACAGCAAAATCCT 1171
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyValArgIleIle----- 98
 DB 1172 AAGTCTCTTTTAAAGTGTGCAAAATTAATGATTGATCTGAGAGAGATCATGATGTC 1231
 QY 98 ----- 98
 DB 1232 ATAGATTAAGAACTAATTCACCTTTGAGATTCATTTGAAGAGATTGAATATATTGCC 1291
 QY 98 ----- 98
 DB 1292 AAGAGCTGATGATCTCCTGAGGAAATATGCTTGTCCATCTCTACTAGATGCTCCAG 1351
 QY 98 ----- 98
 DB 1352 ACTCGCTACAGATAGTGTGATCTCACTCGAATTTTATCCAGTAAGAATGATGTT 1411
 QY 98 ----- 98
 DB 1412 ATGAAAGGAGAGACTGATGAGTGCCTGATTCTGTAGCGCCACTAATATATCTAT 1471
 QY 98 ----- 98
 DB 1472 GAAGAAACAAGACATCTGATTAATATCATGACATCTTTCATGTTTTTCCCAAAGT 1531
 QY 98 ----- 98
 DB 1532 CACGAAGAGAAATTGAGTTTATTTTTCCTCTGAAATGCAAAAAGGTTTCCGTCAATTA 1591
 QY 98 ----- 98
 DB 1592 TACAAATATACATCTATTTTAAAGAAAGCAAAATATAAGCATCAGTGTGGGCTGCT 1651
 QY 98 ----- 98
 DB 1652 GCTCCAGTGAATTTCAAGTGTCTCTCAAGAGAGATAGCAATTACAGTGTGAATGG 1711
 QY 99 ----- ValAspGluValArgArgLeuValTyr 107
 DB 1712 GAAGTCTTGCGCGCATGATCTATATCCAAAGTTGAGAACTCGAAGGCTGGATAT 1771
 QY 108 PheGluGlyThrIlyAspSerProLeuGlnHisIleLeuTyrValValSerTyrValAsn 127
 DB 1772 TTGGAAGGACCAAAAGACTCCCTTTAGAGCATCAGCTGATGATGATGATGATGAT 1831
 QY 128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGln 147

DB 1832 CCTGAGAGGTACCAAGGCTGACTGCTGCTACTCATCTTGTGCAATCACTGAC 1891
 QY 148 HisCysAspPhePheIleSerIlyTyrSerArgGlnLysAsnProHisCysValSerLeu 167
 DB 1892 CACTGTGACTTCTTTTATAGTAAATGATTAACAGAAATCCACACTGTGTCTCTT 1951
 QY 168 TylLysLeuSerSerProGluAspAspProHisCysLysThrLysGluPheTyrAlaThr 187
 DB 1952 TCAAGACTATCAAGTCTCTGAAGATGACCCAACTTGCAAAACAAAGAAATTTTGGCCAC 2011
 QY 188 IleLeuAspSerAlaGlyProLeuProAlaPyrTyrThrProGluIlePheSerPheGlu 207
 DB 2012 ATTTGGATTCAGCAGGTCTCTTCTGACTATACCTCCAAATTTTCTTTTGAA 2071
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 DB 2072 AGTACTACTGATTTTACATTTGATTTGGAGATGCTTACAAAGCTCATATCTACAGCTGGA 2131
 QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGluValGlnLeuValAsnAsn 247
 DB 2132 AAGAAATATCTTACTGTCTGTTCAATATGATGAGTGTCTCAGGTGCAATTGATTAAT 2191
 QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 DB 2192 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTAGGTATATGTGTT 2251
 QY 268 ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
 DB 2252 GTAGCATGACACAGGGAGATCTCTACCCAGGCTTAAATTTAAGGCGCTTTAA 2311
 QY 288 TylLysMetGlyGlnIleGluIleAspAspGlnValGlnGlyLeuGlnTyrLeuAlaSer 307
 DB 2312 TATTAATATGGTCAAAATGAAATTCAGATCAGTGAAGAGACTCAATATCTTACCTCT 2371
 QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyTyr 327
 DB 2372 CGATATGATTTCAATGACTTAAATCTGTGGGCTACCGGCTGCTCATGAGAGATAC 2431
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 DB 2432 CTCTCCCTATGACATTAATGACAGAGTCAAGATATCTTCAAGGTTCTTATGCTGGG 2491
 QY 348 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
 DB 2492 CCAGTCACTGTGAGATCTTCTATGATACAGATACAG 2530
 RESULT 7
 AF221634 3127 bp mRNA linear PRI 05-NOV-2000
 LOCUS
 DEFINITION Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, complete cds.
 AF221634
 ACCESSION
 VERSION AF221634.1 GI:11095187
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS 1 (bases 1 to 3127)
 Abbott,C.A., Yu,D.M., Woolliatt,E., Sutherland,G.R., McCaughan,G.W.
 TITLE Cloning, expression and chromosomal localization of a novel human
 dipeptidyl peptidase (DPP) IV homolog, DPP8
 JOURNAL Bur. J. Biochem. 267 (20), 6140-6150 (2000)
 MEDLINE 11012666
 PUBMED 20467194
 REFERENCE 2 (bases 1 to 3127)
 Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.
 AUTHORS Opposite subunit
 TITLE Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
 Centre, Centenary Institute of Cell Biology and Cancer Medicine,
 Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia
 FEATURES
 Location/Qualifiers

```
source      1. .3127
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /chromosome="15"
            /map="15q22"
            1. .3127
            /gene="DPP8"
            /gene="DPP8"
            /gene="DPP8"
            /note="serine protease; member of prolyl oligopeptidase S9b
            family belonging to enzyme clan SC; alternatively spliced
            forms of this mRNA are found in AF221635, AF221636 and
            AF221637"
            /codon_start=1
            /product="dipeptidyl peptidase 8"
            /protein_id="AAG29766.1"
            /db_xref="GI:11095188"
            /translation="MAAAMETBQLGVEIFETADCENTESODRPLEPPYVERYSMQ
            LKRLADTRKHGYMAKAPHDPMFYKNDPDGPHSDRIYLAISGENSENTLFTSEI
            PKTINRAAVLMLSWKPLDLFOATLDYGMYSREELLRKRKRGVGLASVYHQSG
            TPLFOAGSGIYHVKGQPGQFTQOPLRPNLVETSCNIMRMDPLCPADPDWIAFHSN
            DIWISNIVTREBERLLYVHNELANMEDASAGVATFVIOEEDRSGYWCCEKATTT
            PSQKILRLIVENDSEVEIIVHTSPMLSTRADSEFYKGTANPKYTFPKSEIMI
            DAGRIIDYIDKELLQPEILREGVEVITAAQWTPRGKXAMGILLDRQTLQIVLIS
            PLFTIPEVDVNERKRLISVPSVPLTIYETTDIMINIDIFHVFPQSHBEIEF
            IPASECKTGRHLYKITSILKESKYRSSGGLPAPDFEYCPKEIEAITSGEHEVIGR
            HGSNIQVDEVRLLVPEGTKDPLEHHLVVSIVNDEYTRLDREYASHSCISQCHD
            FFSKSNQKNPCHVSLYKLSPEDDPTCKEFMATILDSAGLPEDYTPPELFSFS
            VTGFTLGMVLRHDPQPKKPTVLFTIGQPOVLNNRPFQVKEFRNLNTLASIGYV
            VVVDNRGSCHRGLKREGAPKYKQIOELIDDOVEGICVLAIRYDFIDLRVGHGMSY
            GGTLISLMALMQRSDIRFVVAIAGAPVTLMFTFYDTGYTERMGHPDQEQYIGYSAMQ
            AEKPESEPRRLILHGFIDENVAHFATSTILSLVRAKPYDLQIYPOHRSHIRVES
            GRHYELHLHYIOLNLSRIALXVI"

BASE COUNT      943 a      637 c      706 g      841 t
ORIGIN

Alignment Scores:
Pred. No.:      1,67e-180      Length:      3127
Score:          1836.50         Matches:      360
Percent Similarity: 67.54%      Conservative: 0
Best Local Similarity: 67.54%      Mismatches: 0
Query Match:      95.01%         Indels:      173
DB:              9              Gaps:      1

US-10-070-464-7 (1-360) x AF221634 (1-3127)

QY      1  GIUGLUASPAIAARSERALAGLYVALAATHRPHVAILEUNGINGUINUPHEASPARQ 20
      DB 943 GAAGAAGATGCCGATCAGCTGAGTCGCTACTTTGTTCTCAAGAAGATTGATAGA 1002
QY      21  TYRSEGLYTYRTTPCYSPOLYVALAGLINTHTRPROSEGLYGLYLSILEU 40
      DB 1003 TATTCGGCTATGCGTGTCCTCAAAAGCTGAACAACCTCCAGTGTGTAAATCTT 1062
QY      41  ARGILEUITYGJUGLUNASPAJUSERGLUVALGINTLEHISVALITRRSEPRQ 60
      DB 1063 AGAATTCATATAAAGAAATGATGATTCGAGGTGAATTTATCATGTTACATCCCT 1122
QY      61  METLEUGLINTHARGTALAAESPSEPHEARGYTYRPROLYSTHRGLYTHRALAASPRO 80
      DB 1123 ATGTTGAAAACAAAGAGGGCAGATTCATTCGTTATCCTTAAACAGGTACAGCAAAATCCT 1182
QY      81  LYSVALITHRPHLYSNETSERGLUILEMETILEASPAJAGLUGLYARGILLE----- 98
      DB 1183 AAAGTCACTTTAAAGATGTCAGAAATTAATGATGATGAGAAAGATCATATGATGTC 1242
QY      98  ----- 98
      DB 1243 ATAGATAAGAACTAATTCACCTTTTGAGATTCATATTGAAAGAGTTGAATATATTGCC 1302
QY      98  ----- 98
```

```
DB      1303 AGAGCTGATGAGACTCTCGAGGGAAAATATGCTTGCTCATCTACTAGATGCTCCAG 1362
QY      98  ----- 98
DB      1363 ACTCGCCCTACAGATAGTGTGTGATCTCACTGAATATTATATCCAGTAGAAGATGATT 1422
QY      98  ----- 98
DB      1423 ATGAAAGGACAGACTCATTTAGTCAGTCACTGATTCTGTGACGCCACTAATTATCTAT 1482
QY      98  ----- 98
DB      1483 GAAGAAACACAGACATCTGATTAATATCATGACATCTTTCATGTGTTTCCCAAGT 1542
QY      98  ----- 98
DB      1543 CAGAAAGAAATTTAGTTTATTTTTCCTGTGATGCAAAAACAGTTTCCGTCAATTA 1602
QY      98  ----- 98
DB      1603 TACAAATTAATCATTTATTTAAAGAAAGCAAAATATAACGATCCAGTGGGCTGCCT 1662
QY      98  ----- 98
DB      1663 GCTCCAGTGAATTTCAAGTGTCTTATCAAGAGAGATAGCAATTACCAAGTGGTGAATGG 1722
QY      99  -----ValSPGLUVALHARGALEUValTYR 107
DB      1723 GAAGTCTTGCCCGGATGATCTAATATCAAGTTGATGAAGTCAAGAGGCTGGATATAT 1782
QY      108 PHEGLUGLYTHRLYSASPERPROLENGIUNHISHILEUTYRVALIASERTYRVALASN 127
DB      1783 YTTGAAGGACCAAAAGACTCCCTTTAGAGCATCACCTGATGATCTACGTTACGTAAT 1842
QY      128 PROGLUGLVALTHRARGLEUTHIRASPARAGLYTYRSETHISERCYSCYSLISERGIN 147
DB      1843 CCTGAGAGAGTGACAAAGCTGACTGACCGTGCTACTCATCTTGCTGTCATCAAGTCAG 1902
QY      148 HISYASPPHEPHEILLESERTYRSETHISERGINLYASNPROIHISCYVALISERLEU 167
DB      1903 CACTGGAATCTTTTAAAGTAAAGTATGATTAACCAAGAAATCCACACTGTGTGCTCTT 1962
QY      168 TYRILEUUSERSEPROGLUNASPARPROTHRCYSLVETHRLYEGJUPHETPALATHR 187
DB      1963 TACAAAGCTATCAAGTCTGAAGATGACCCAACTGCAAAACAAAGAAATTTTGAGCCACC 2022
QY      188 ILELEUASPERALAGLYPROLEUPROASPTYRTHRPROPROGLUINLEPHESEPRHEGU 207
DB      2023 ATTTTGATTTACAGAGGTCTCTTCTGACTATCTCCCAAGAAATTTTCTTTTGAA 2082
QY      208 SEETHRTHGLYPHETHRILEUTYRGIMETLEUTYRLYSPROHISASPLEUNGINPROGLY 227
DB      2083 AGTACTACACTGATTTTACATTTGATGGATGCTCTACMAOCCCTCATATCTTACAGCCTGA 2142
QY      228 LYSVSTYRPROTHRALAUPHEILLEYRGJYGLYPROGLUVALGINLEUVALAASN 247
DB      2143 AAGAAATATCCACTGCTGCTGTCATATATGTTGTCCTCAGGTGAGTTGGGAATAT 2202
QY      248 ARGPHLYSGLYVALIYSYRPHARGLEUASNTHRLLEUALASERLEUGLYTYRVAL 267
DB      2203 CGGTTTAAAGAGTCAAGATATTCCGCTTGAAATACCTTGAAGCTCTCTAGGTTATGCGTT 2262
QY      268 VALVALIILEASPARHARGLYSERCYSHISARGLYLEULYRPHGLUGLYVALAPHELYS 287
DB      2263 GTAAGTATGACAAACAGGGGAACTCGTGCACCAAGGGCTTTAAATTTGAAAGCGCTTTTAA 2322
QY      288 TYRLYSMERGJYGLNILEGJUILEASPARGINVALIGJYLEUNGINTYRLEUALASER 307
DB      2323 TATTAATGGGTCAAAATTAATAATTGACATCAGGTGGAAGACTCCAAATATCTAGCTTCT 2382
QY      308 ARGTYRASPPHEILEASPLEUASPARGYVALIGYIILEHISGLYTYRSETYRGLYGLYTYR 327
DB      2383 CGATATGATTTCAATGACTTAGATCGTGTGGGCAATCCAGGCTGCTATGAGAGATAC 2442
```


QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
DB 2443 CTCTCCCTGATGGCATTATGACAGAGCTCAGATATCTTCAGGGTCTCTATTGCTGGGCC 2502
QY 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
DB 2503 CCAGTCACCTCTGTGATCTTCTATGATACAGATACAGC 2541
RESULT 8
AX354793
LOCUS AX354793 3143 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 1 from Patent WO01/9473.
ACCESSION AX354793
VERSION AX354793.1 GI:18619526
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Meyers, R.A. and Williamson, M.
AUTHORS 21953, a human prollyl oligopeptidase family member and uses thereof
JOURNAL Patent: WO 01/9473-A 1 25-OCT-2001
Millemmium Pharmaceuticals, Inc. (US)
FEATURES
source
1. 3143
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
229..2877
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD22990.1"
/db_xref="GI:18619527"
/translation="MAAMETBQLGVEIFETADCEINISODPKLEPPYVERYSNQ
LKKLADTRKHYGMMAKAPHDPMFVKNDPDGPHSDRIYILAMSENEENLTFSEI
PKTINAAVLMLSMKPLDLFOATLDYGYSRBELRLRRKRGVIGIASYTHYQSG
TFPQAGSGIYHKDQGPQFTQOPLRPMLVETSCNIMDEPLCPADPMIAFISN
DIWISNITREERRLTYVHNELANMEDASAVATFVLOEEDRSQWMCCKATF
PBGKRLIYVEENDESEVEIITHVSPMLETRADSRYPKGTANPKYTPKSRIM
DABGRITIDIDKELIOPFEILPBGVEYIARAGTPBGRKAMSLIDRSQTRQIVIS
PELFIPEDDVNERQLIESVPSVPLIIEETIDMINIHIFHVPQSHBEIEIF
IPASECKTPRRLYKITSILKSKYSSGGAPAPDFKCIKEELAIISGEWELGR
HGSNIOVDREVRILVYEGTDSPLLEHLVVSYNVEGTPLRDLRGVSHSCISOHD
FLISKSNKNPCHVCVLYKLSPEDDPTCKTEKFMATILDSAGPLDYTPPELFSFS
TTGFTLYGMLYKPHDLPQKKYPTVLFIYGPOVOLNRRPKGVYERLNTLSAGY
VYVIDRSGCHRGKREGAFKYMGOIILIDDOYEGIOYLASRPDIFDLRVGHWGSY
GGTSLMALMGRSDIRPRVALIAGAPVTLMIPTDGTERTERMGHPDQNGCYTIGSVAMQ
AEKFPSEPNRLILHGFDENVHFAFTSILSLFLVNAKPYDLQIYPOBRHSIRVES
GHEYELHLHYOENLGRISIAALKVI"
BASE COUNT 943 a 644 c 712 g 844 t
ORIGIN
Alignment Scores:
Pred. No.: 1,68e-180 Length: 3143
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 6 Gaps: 1
US-10-070-464-7 (1-360) x AX354793 (1-3143)
QY 1 GUGLUASPAAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
DB 958 GAAAGAGATGCCGAGATCAGCTGAGTGGTCTTCTTCTTCCAAAGAAATTGATAGA 1017
QY 21 TyrSerGlyTyrTrpTyrCysProIysAlaGlnThrThrProSerGlyGlyIleLeu 40
DB 1018 TATTCTGCTATTGTGGTGTCCAAAGCTGAAACAACTCCAGTGGTGTAAATTCCT 1077
QY 41 ArgIleLeuTyrGlnGluLusAsnAspGluSerGluValGlnIleIleHisValThrSerPro 60

DB 1078 AGAATTCATATAGAAAGAAATATGATGATGAGGGAATTAATTCATCTCCCT 1137
QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProIysThrGlyThrAlaAspPro 80
DB 1138 ATGTGGAAACAAAGAGGAGCAATTCATTCCTTATCCAAACAGTACAGCAATTCCT 1197
QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGlnIlyArgIleIle 98
DB 1198 AAGTCACCTTTAAGATGTCAGAAATATGATGATGATGATGATGATGATGATGATG 1257
QY 98 98
DB 1258 ATAGATAGAGAACTAATTCACCTTTTGAATTCATTTGAAGAGTGAATATATTGCC 1317
QY 98 98
DB 1318 AGAGCTGATGAGTCTCTGAGGAAATATGCTTGGTCACTCTACATGATGCTCCAG 1377
QY 88 98
DB 1378 ACTGCGCTGAGATAGTGTGATCTCAGCTGAATTTATATCCAGTAAAGATGATGTT 1437
QY 98 98
DB 1438 ATGGAAGCAGAGACTCATTTGATGATGATGATGATGATGATGATGATGATGAT 1497
QY 98 98
DB 1498 GAAAGAAACAAACAGATCTGATTAATATTCATGATCTTTATGTTTCCCAAGT 1557
QY 98 98
DB 1558 CACGAAGAGAAATGATTTATTTTGTGCTGATGCAAAAAGAGTTTCCGTCAATTA 1617
QY 98 98
DB 1618 TACAAATTAATCATCTATTTTAAAGAAAGCAATATTAAGATCAGTGGGCTGCT 1677
QY 98 98
DB 1678 GCTCCAGGATTTCAAGTGTCTTATCAAAAGAGAGATAGCAATTACAGTGGTGAATGG 1737
QY 99 98
DB 1738 GAAATTCCTGGCGGATGATGATATATCAAGATGATGATGATGATGATGATGAT 1797
QY 108 PheGluGlyThrLysAspSerProLeuGlnHisIleLeuTyrValIleSerTyrValAsn 127
DB 1798 TTTGAAGGACCAAAAGATCCCTTTAAGAGCATCACCTGTACGTACGTACGTAAAT 1857
QY 128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGln 147
DB 1858 CCTGAGAGAGTACAAAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1917
QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
DB 1918 CACTGAGCATCTTTTATAGTATGATGATGATGATGATGATGATGATGATGATGAT 1977
QY 168 TyrIleLeuSerSerProGluAspAspProThrCysIleThrLysGluIlePheIleThr 187
DB 1978 TACAAAGCTATCAAGTCTTCAAGATGACCAACTTGCAAAACAAAGAAATTTGGGCAAC 2037
QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
DB 2038 ATTTTGATTTACAGAGTCTCTCTCTGATACATACCTCCAGAAATTTCTTTTGA 2097
QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
DB 2098 AGTACTACGAGATTTCATTTGATGATGATGATGATGATGATGATGATGATGATG 2157
QY 228 LysIleTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247

Db 2898 GAAGAAACAACAGACATCTGGATMAATCATCATCTTTCATGTTTTTCCCAAGT 2957
 QY 98 ----- 98
 Db 2958 CACGAAGAGAAATTGAGTTTATTTTGGCTCTGAATGCAGAAAGGTTTCCGTATT 3017
 QY 98 ----- 98
 Db 3018 TACAAATTACATCTATTTTAAAGAAAGCAATATAAAGATCCAGTGGTGGCTGCT 3077
 QY 98 ----- 98
 Db 3078 GCTCCAGTGAATTCAGTGTCTCATTAAGAGAGATAGCAATTCACAGTGGTGAATG 3137
 QY 99 ----- ValAspGluValArgArgLeuValTyr 107
 Db 3138 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAAGTGAATGAGTCAAGAGGCTGATAT 3197
 QY 108 PheGluGlyThrLysAspSerProLeuGluHisLeuTyrValValSerTyrValAsn 127
 Db 3198 TTGGAGGACCCAAAGACTCCCTTTAGAGCATCAGCTGACCTAGCTTACGTTACGTTAA 3257
 QY 128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
 Db 3258 CCTGGAGAGGTGCAAGAGCTGAGTGAACCGTGGCTACTCATTTCTGTGCATCAGTCA 3317
 QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAspProHisCysValSerLeu 167
 Db 3318 CACTGTGACTTCTTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 3377
 QY 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
 Db 3378 TACAACTATCAAGTCTCGAAGATGACCCCACTTGCAAAACAAGGATTTTGGGCAC 3437
 QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
 Db 3438 ATTTGGATTCAGACAGTCTCTCTCGTACTATACCTCCAGAAATTTTCTTTTGA 3497
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 Db 3498 AGTACTACTGATTTTACATTGATGAGATGCTTACAGCCTCATATCTACAGCCTGGA 3557
 QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 247
 Db 3558 AAGAAATATCTACTGCTGCTGATCATATATGAGTGTCTCAGGTGCAAGTTGGATTAAT 3617
 QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 Db 3618 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTAGGTTATGAGTT 3677
 QY 268 ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGluValAlaPheLys 287
 Db 3678 GTAGTATAGCAACAGGGGATCTGTCCAGCGGCTTAAATTTGAAGCGCCTTTAAA 3737
 QY 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
 Db 3738 TATAAATGGTCAATATAGAAATTCAGATCAGGTGGAAGGACTCCAAATATCTAGCTTCT 3797
 QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyThrSerTyrGlyGlyTyr 327
 Db 3798 CGATATGATTTTCAATTCATTAAGTGTGAGGATCCAGCGCTGCTCATGAGGATATC 3857
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 Db 3858 CTCTCCCTGATGGCATTAATGACAGAGTCAAGATATCTTACAGGTTCTATGCTGGGCG 3917
 QY 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
 Db 3918 CCAATCACTCTGTGATCTTCTATGATATAGATACAGATACAG 3956
 RESULT 10
 AX608735

LOCUS AX608735 4829 bp DNA linear PART 17-FEB-2003
 DEFINITION Sequence 12 from Patent WO0231134.
 ACCESSION AX608735
 VERSION AX608735.1 GI:28404303
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 61 Qi S., Akinsanya, K.O., Riviere, P.J. and Junien, J.L.
 AUTHORS Novel serine protease genes related to dppiv
 TITLE Patent: WO 0231134-A 12 18-APR-2002;
 JOURNAL Ferring BV (NL)
 FEATURES
 source Location/Qualifiers
 1..4829
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 1466 a 886 c 1017 g 1460 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,41e-178 Length: 4829
 Score: 1816.50 Matches: 360
 Percent Similarity: 67.29% Conservative: 0
 Best Local Similarity: 67.29% Mismatches: 0
 Query Match: 93.97% Indels: 175
 DB: 6 Gaps: 1
 US-10-070-464-7 (1-360) x AX608735 (1-4829)
 QY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluIupheAspArg 20
 Db 943 GAAGAGATGCCAGATCAAGCTGAGTGGCTACCTTTGTTCTCCAGAGAAATTTGATAGA 1002
 QY 21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyLysIleLeu 40
 Db 1003 TATTCGGCTATGTTGGTGGTGTCCAAAGCTGAAACAACCTCCAGTGGTGAATATCTT 1062
 QY 41 ArgIleLeuTyrGluGluAspGluSerGluValGluIleHisValThrSerPro 60
 Db 1063 AGAATTCATATGAAAGAAATGATGATGAGTGAATTTTCAATGTTACATCCCT 1122
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAspPro 80
 Db 1123 ATGTTGAAACAAAGAGGCGGCAATTCCTGTTATCTTAAACAGGTACAGCAATCTCT 1182
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluIleIle----- 98
 Db 1183 AAAGTCATTTTAAAGATGTCAGAAATTAATGATGCTGAAGAGGATCATGATGTC 1242
 QY 98 ----- 98
 Db 1243 ATAGATTAAGAACTAATTCACCTTTTGAGATTCTATTGAAAGAGTGAATATATTGCC 1302
 QY 98 ----- 98
 Db 1303 AGAGCTGATGAGACTCTTGAAGGAAATATGCTGTGTCATCTACTAGATCGCTCCAG 1362
 QY 98 ----- 98
 Db 1363 ACTGCGCTACAGATAGTGTTCATCTCAGCTGAATTTATATCCAGTAGAAGATGATGTT 1422
 QY 98 ----- 98
 Db 1423 ATGAAAGGACAGACTCATTTAGTCAGTGGCTGATTCTGTGACGCCACTAATTATCTAT 1482
 QY 98 ----- 98
 Db 1483 GAAGAAACAACAGACATCTGATTAATATCCATGACATCTTTCATGTTTTTCCCAAGT 1542
 QY 98 ----- 98

```

Db      1543 CACGAAGAGGAATTGAGTTATTTTGGCTCTGAATGCAGGTTTCCGTCAATTTA 1602
QY      98 ----- 98
Db      1603 TACAAATTACATCTATTTTAAAGAGCAAAATATTAACGATCCAGTGGTGGCTGCC 1662
QY      98 ----- 98
Db      1663 GCTCCAGTGAATTCAGTGTCTCTATCAAGAGAGATAGCAATTAACAGTGGTGAATG 1722
QY      99 ----- ValAspGluValArgArgLeuValTyr 107
Db      1723 GAAGTTCTGGCCGGCATGATCTAATATCCAAAGTTGATGATCAAGAGCTGGATATAT 1782
QY      108 PheGluGlyThrLysAspSerProLeuGluHisLysLeuTyrValValSerTyrValAsn 127
Db      1783 TTGGAAGGACACCAAAACATCCCTTTAGAGCATCACCTGACCTAGCTAGTACGTAAT 1842
QY      128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysLysSerGln 147
Db      1843 CCTGGAGAGGTGACAAAGCTGACGACCGTGGCTACCATCTTGTGTCATCAGTCAAG 1902
QY      148 HisCysAspPhePheHisSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db      1903 CACTGAGACTTCTTTAAGTAGTAGTAGTAGTAACCAAGAGATCCACACTGTGTGCTCT 1962
QY      168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPalaThr 187
Db      1963 TACAGCTATCAAGCTCTGAAAGTAGACCCCAACTTGCAAAACAAAGGATTTTGGGCCAC 2022
QY      188 IleLeuAspSerSerIleGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db      2023 ATTTGGATTCACAGAGTCTCTCTCTGACTATACCTCTCAAAATTTCTCTTTTGA 2082
QY      208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db      2083 AGTACTACTGAGATTTCATTGATGGATGCTCTTCAAGCTCATGATCTACAGCTCGA 2142
QY      228 LysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnLeuValAsn 247
Db      2143 AAAAAATATCTCTACTGCTGCTGCATATATGAGTGCTCTCTCAGGCTGAGTGTGAATA 2202
QY      247 snArpPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValY 267
Db      2203 ATGGGTTTAAAGAGTCAAGTATTTCCGCTTGATACCTTACGCTCTCTAGGTTATGTG 2262
QY      267 aValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyValaPhe 287
Db      2263 TTGTAGATAGACAAACAGGGGATCTGTCACAGGGGCTTAATTTGAAGCGCCTTTA 2322
QY      287 YstYrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 307
Db      2323 AATATATAAGGGTCAAAATGGAATGACGATGAGGAGAGGACTCCCAATATCTAGCTT 2382
QY      307 eArGlyTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPseTyrGlyGlyT 327
Db      2383 CTGCAATATGATTCATGACTTAGATCGTGTGGCAATCCAGGCTGGTCTATGGAGGAT 2442
QY      327 YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 347
Db      2443 ACCTCTCCCTGATGGATTAATGACAGGCTCAGATATCTTCAGAGTTGCTATTTGGGG 2502
QY      347 laProValThrLeuTyrPhePheTyrAspPheTyrThr 360
Db      2503 CCCAGTCACTGTGTGATCTTATGATACAGAGATACAG 2543

```

```

RESULT 11
BC043124      4799 bp      mRNA      linear      ROD 22-APR-2003
LOCUS      Mus musculus dipeptidylpeptidase 8, mRNA (cDNA clone MGC:58057
DEFINITION      IMAGE:6410075), complete cds.
ACCESSION      BC043124

```

```

VERSION      BC043124.1 GI:27695449
KEYWORDS      MGC.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      1 (bases 1 to 4799)
AUTHORS      Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, U., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Frange, C., Rana, S.S., Loggellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
Mckernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Schechenko, Y.,
Boutard, G.G., Blakesley, R.W., Touchman, D.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Gutterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
TITLE      JOURNAL
MEDLINE      22388257
PUBMED      12477932
REFERENCE      2 (bases 1 to 4799)
AUTHORS      Strausberg, R.
DIRECT SUBMISSION
SUBMITTED (09-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas U. Casavant.
Web site: http://genome.uiowa.edu
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
Fisher, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
Scheetz, T., Smith, C., Smir, E., Tack, D., Trout, K., Walters, J.,
Casavant, T., Soares, M.B.
Clome distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.lnl.gov
Series: Plate: Row: Column: 0.
Location/Qualifiers
1..4799
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:58057 IMAGE:6410075"
/tissue_type="Brain, enriched mouse brain 12.5dp"
/clone_lib="NIH BMAP_F00"
/lab_host="DHIOB"
/note="Vector: pYX-ASC"
1..4799
/gene="Dpp8"
/db_xref="LocusID:74388"
/db_xref="MGI:1921638"
288..2966
/codon_start=1
/product="Dpp8 protein"

```

```

/protein_id="AAH43124.1"
/db_xref="GI:27695450"
/db_xref="LOCUSID:74388"
/translation="MKIPSGRCNMAAAMETEOIAGVIEFETACBCEGNGESODRPLEP
FYVERYSMSOLKLIADTRKXGYMAAPHDPMFVMDPDGPDGSDRYVYLAMGEN
RENTLFYSSEIPKTRINRAVIMLSMKPLIDROATIDYMGVSEBELERKIGIVGI
AAVDYHPSGTFLEFGAGSITHIKDGHGFTQPLRPLVTSOPNIRMDKCPAD
PDMIAFHSNDMISNLVTRERITVYELNANMEBRSAGVATFVLOEPDYSK
YVPCPOARSPSGKILRLILEENDESEVEIIHVTSPMLETRADSPRYPKGTANPK
VTFKSEIIVDAAGGIIDVIDKEILOPEILEGVEYIARAGMTEGHKMSILDRS
OTHOIVLISPELFIPEVDAMDRORLIESVPSDVPLIYEETDIDMINDIHVF
POTHEDELEFIASCKTGPRPLIKTISLKSRYKRSGLPASPDKCPKEKITI
TSCEMEVCRHESNIVDPAKLVFEGTKDSLEHHLVTSYANPGBVRLTDGYS
HSCCLSRHCDPFIISKYSNQKNCVSLKLSPEDDPVHKKEFWATLIDSAGLPDY
TPPELFESFESTGFTLYGMLYKPHDLOPKKPYVLFIYGGVQLVNNRFGVYFR
INTLASIYGVVVVINDRGSCHRGKLEFGAFKYMGOIHIIDDOVEGLQYIASOYDIDL
DRVGHMSYGGYSLMALMORSIDIFRAVIGA PVTLMIFDTGTGYEYEMGHQDNEQ
GYLGSVAMQAEKRPSPERNLLIHGPIIDENHPAHTSILSLFLVRAKGPDLQIYPO
ERHSIRVPSGSEHRLHLDHYIQENLGSRIALAKVI"

```

BASE COUNT 1373 a 1015 c 1103 g 1308 t

ALIGNMENT SCORES:

Pred. No.: 5,816-172 Length: 4799
 Score: 1756.50 Matches: 343
 Percent Similarity: 65.67% Conservative: 7
 Best Local Similarity: 64.35% Mismatches: 10
 Query Match: 90.87% Indels: 173
 DB: 10 Gaps: 1

US-10-070-464-7 (1-360) x BC043124 (1-4799)

```

QY 1 GIUJASPAALARGSERALAGLYVALATHRPHEVALLEUGLUGLUPHEASPARG 20
DB 1047 GAAGAGGATCCAGATCAGCTGCTGTGGCCACTTGTCTTCAAGAAATTTACACGA 1106
QY 21 TYRSERGLYTYRTTTPCYSPROLYSALAGLUTHRTPROSEGLYGLYLYSILLEU 40
DB 1107 TACTGTGCTACTGTGTGTGTCGCCAAGCGGAAAGACTCTGATGCTGATTAATTCCT 1166
QY 41 ARGILEUITYRGLUGLUNASPSGLUSERGLI VALIGLILLEIHSIVALTHTSERPRO 60
DB 1167 AGAATTCCTCTAAGAAATAATGATGATGTGAGTGAGATTAATTCATGTCGCCCC 1226
QY 61 METLEUGLUTHRARGALASPSERPHARGIYRPROLYSTHRLYTHRLAASPRO 80
DB 1227 ATGTGGAAACAGAGGAGGACATTCCTTCGTATCCCAAAACAGGACAGCAAAACCA 1286
QY 81 LYSVALTHRPHELYMETSERGLIULEMETILLEAPHALAGLUGLYARGILLE----- 98
DB 1287 AAGTCACCTTCAAGATGTCGAGATGTGTGTGATGCTGCAGAGGAGGATTTAGATGTC 1346
QY 98 ----- 98
DB 1347 ATAGATMAAGAACTGGTTCAACCTTTCAGATTCCTGTTGAGGAGATTGATATTTGCC 1406
QY 98 ----- 98
DB 1407 AGAGCCGATGACTCCAGAGGAAACATGCTGTCATCTCTACTAGACCGTTCACG 1466
QY 98 ----- 98
DB 1467 ACTCACTGCAGATAGTCTGATCTCCCTGAGTTATTCATCCAGTAGAAGATGATGCC 1526
QY 98 ----- 98
DB 1527 ATGCACAGACAGACTTATAGAGTCAGTCTCTGACTCTGTACACCACTGATCATCTAT 1586
QY 98 ----- 98
DB 1587 GAAGAAACAGACAGCATCTGATATAATATCCAGATATTTTCATGTTTTCCTCAACT 1646
QY 98 ----- 98

```

```

DB 1647 CATGAAGATGAATTAAGATTATTTTTCCTTCGTAATGCMAAACAGGTTTTCGTCATCTG 1706
QY 98 ----- 98
DB 1707 TATTAATACATCATCTATTTTAAGAGAGCAATATTAACGGTCCAGTGTGACTACT 1766
QY 98 ----- 98
DB 1767 GCCCAAGTATTCAAGTGTCTATCAAGAAAGAAATTAACATTACAGAGTGATGATG 1826
QY 99 ----- 98
DB 1827 GAAGTACTTGCGCGCATGATCTATATCTGCGTTGATGAAGCCAGAAAGCTGGGTAC 1886
QY 1887 TTTGAAGGACCAAGACCTCTCTTGGAGCATCCGTATGATGACCAAGTATGCAAC 1946
DB 1947 CCTGAGAAAGTGTGAGGCTGACGACCGTGGCTACTACACTCCGCTGCTGACGCGG 2006
QY 1948 HISCYASAPPHAEHELSESLYSTRSERANGILYASRPROHISCYVALSERLEU 167
DB 2007 CATTTGACTCTTCTATAGTAAGTACAGCAACGAAAGAAATCCACTGTGTCTTC 2066
QY 168 TYRYSLEUSERSEPROGLUNASPSAPROTHRCYSLSLTHLYSGLUPHETPALATHR 187
DB 2067 TACAACTCTCAAGCTCTGAGGATGACCCAGTTCATAAACAAGAAATTTGGGCCACC 2126
QY 188 ILEUASPSERALAGLYPROLEUPROASPTYRTHRPROPROGLIULEPHESERPHGLU 207
DB 2127 ATTTGGATTGAGAGGCTCTCTCTGACATCAACCCCTCAAGAAATTTTCTTTGAA 2186
QY 208 SERTHTRGLYPHETHLEUTHRGLYMETLEUTHRYSLPHROHISASPLEUGLUPROGLY 227
DB 2187 AGTACTCTGATTTTACACTGATGAAATGTTGTATTAACCTCAAGACCTTACAACTGGA 2246
QY 228 LYSLSYTYRPROTHVALLEUPHELELYRGLYGLYLPROGLI VALIGLILEUVALAENAN 247
DB 2247 AAGAAATACCCACTGTGTTTATTCATATATGAGTGTCCTCCAGTGCAGCTGTGAACAT 2306
QY 248 ARGPHLYSGLYVALIYSTRPHEARGLEUASNTHRLLEUALSERLEUGLYTYRVALVAL 267
DB 2307 CGGTTAAAGAGTCAAGATATTCGCGCTGAACACCTGCGCTCCCTGGGTTATGCGTT 2366
QY 268 VALVALILEASPSAHLARGLYSERCYSHISATGGLYLEULYPHEGLUGLYALAPHELYS 287
DB 2367 GTGGATGATGACAAACAGGGGATCCCTGTCAACGAGACTTAAATTTGAAGCGCCCTTTAA 2426
QY 288 TYRYSMETGLYGLIILEGLIULEASPSAGLIVAGLUGLYLEUGLITHTYRLEUALASER 307
DB 2427 TATTAATGCGTCAATATGAAATGATATCAAGTGAAGAGCTCCAGTACTAGCATCT 2486
QY 308 ARGTYRASPHEHLEASPLEUASPARVALIGLYLEHISGLYTRPSEYRGLYLYTYR 327
DB 2487 CAGTATGACTTCAATGACTTGATGAGAGGACATCCACGCTGTGCTATGCTGCTAC 2546
QY 328 LEUSERLEUMETALILEUMETLARGSERAPHLEPHARGVALALALALAGLYALA 347
DB 2547 CTCTCCCTGATGCGATTAATGACAGGTCGATATCTCCGGGTGCTATGCTCGGGGCC 2606
QY 348 PROVALTHRLEUTHRILEPHETRYSPHTRGLYTYRTHR 360
DB 2607 CCAATCACCTGTGATCTTCTATGATACAGGATACAG 2645

```

RESULT 12
 BC030688 3130 bp mRNA linear PRI 30-MAY-2002
 LOCUS Homo sapiens, similar to dipeptidyl peptidase 8, clone MGC:26191
 DEFINITION IMAGE:4822550, mRNA, complete cds.
 ACCESSION BC030688

VERSION BC030688.1 GI:21265132
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 3130)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAY-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAC Plate: 33 Row: d Column: 5
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 18450277.
 FEATURES
 source
 Location/Qualifiers
 1..3130
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:26191 IMAGE:4822550"
 /tissue_type="Testis"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Vector: pBluescript"
 372..2867
 /codon_start=1
 /product="Similar to dipeptidyl peptidase 8"
 /protein_id="AAH0688.1"
 /db_xref="GI:21265133"
 /translation="MAAMTEQLGVEIFPTADCEENIESQDRPKLEPFYVERYSMSQ
 LKTLADTRKYHGMMAKAPHDPMFVRNDPDPHSDRIYILMSGSENRNTLFSYBI
 PRTINRAAVLMLSWKPLDLFOATLDYGMTSREELIREKRIGTVAGISDYHOGSG
 TFLFOAGSGIYHVGKGGPOGFTOOPLPBNLVETSCPIRMIDPLCPADPMIAFISN
 DIMISNIVREERRLTYVHNEILANMEDASAGVATFVCOEEDRSGVWMCRAETT
 PSGKILIRITRENDSEVEIHYTSPMLTBRASADSRVRYKGTANPKYTFKSEIMT
 DAGRIITIDVDELIOPELTFEGVEYIAAGWTPROKTAWSILIDRSQRLQIVLIS
 PELFIPVEDVMERQRLIESVPSVPLIYEBETDIIWIHDI FHFVPSHEEIEF
 IPASECTQFRHLKYKITSILKESKYKSSGGLPAPSPKPIKEEIALITSGEWEVLGR
 HGSNIQVDEVRRLVYFEGTKDSPLEHLIYVSYVNPGRVRLDRGSHSCCSOHD
 FPI SKSNQGNPHCVSLYKLSPEDDPTCKTFEMATILISAGPLPYTPPELFSFS
 TTGFTLYGMLYKHGDLQPKKYPYVLIYGGQVQVLYNNEPKVKKYERLNTASLGV
 VVITDRGSCHRGLKTEGAKYKVAIALAGPVITLMTFTDGTERTYHKGHPDQBGY
 LGSVVAQAEKFPSEPNKLLHGLFDENVFATSTILSFLVAGKPYDILQIYPERH
 SIRVPSGEHYELHLHYIOENLGSRIALAKVI"
 CDS
 BASE COUNT 955 a 626 c 717 g 832 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.55e-149 Length: 3130
 Score: 1534.00 Matches: 309
 Percent Similarity: 57.97% Conservative: 0
 Best Local Similarity: 57.97% Mismatches: 0
 Query Match: 79.36% Indels: 224
 DB: 9 Gaps: 2
 US-10-070-464-7 (1-360) x BC030688 (1-3130)

QY 1 GIUIGUASPALAArSeraIaGlyValAlaThrPheValLeuGluGluGluPheAspArg 20
 DB 1101 GAAGAGAGATCCAGATCAGCTGAGTGGTACCTTTGTTCTCAAGAAAGAAATTTGATPAG 1160
 QY 21 TYSERGLYThrTrpCysProlysaIaGluThrThrProSerGIyGlyysIleLeu 40
 DB 1161 TATTCGCGTATTTGGTGGTGTCCAAAGCTGAACACACTCCAGTGGTGTAAATTCCT 1220
 QY 41 ArgIleLeuTyrgluGluuAspGluSerGluValaGluIleIleHisValThrSerPro 60
 DB 1221 AGAATTCATATGAAAGAAATATGATGATCTGAGTGAATATATTCATGTTACATCCCT 1280
 QY 61 MetIleGluThrArgArgValaAspSerPheArgTyrrProlysthrGlyThrAlaAspPro 80
 DB 1281 ATGTTGGAAACAAAGAGGGGCAATTCCTGTTACCTPAAACAGGTACAGCAAAATCCT 1340
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle 98
 DB 1341 AAAGTCACTTTTAAAGATGTCAGAAATTAATGATTGATGCTGAAGAAAGATCATGATGTC 1400
 QY 98 ----- 98
 DB 1401 ATGATTAAGAACTAATTCACCTTTGAGATTCATTTGAAGAGTTGAATATATATGCGC 1460
 QY 98 ----- 98
 DB 1461 AGAGCTGATGATCCTCTGAGGAAATATGCTTGGTGCATCTACTAGATCGCTCCAG 1520
 QY 98 ----- 98
 DB 1521 ACTGCGCTACAGATAGTGTGATCTGACCTGAATTTATATCCAGAGAGATGATGTT 1580
 QY 98 ----- 98
 DB 1581 ATGAAAGCAGAGACTCATATGATGATGCTGATTCGTGTGACGCCACTAATATATCTAT 1640
 QY 98 ----- 98
 DB 1641 GAAGAAACAACAGACATCTGAGTAATATCCATGACATCTTGATGTTTCCCAAGT 1700
 QY 98 ----- 98
 DB 1701 CACGAAGAGAAATGAGTTTATTTTGGCTCTGAATGCAAAACAGGTTCCGTCATTTA 1760
 QY 98 ----- 98
 DB 1761 TACAAATTAATCATCTATTTTAAAGAAAGCAATATAACGATCCAGTGGGCTGCT 1820
 QY 98 ----- 98
 DB 1829 GCTCCAAAGTATTTCAAGTGTCTCTATCAAGAGAGATAGCAATTACCACTGGTGAATGG 1880
 QY 99 ----- ValAspGluValArgArgLeuValTyr 107
 DB 1881 GAGGTTCTTGCGCGCATGATGATTAATATCAAGTATGATGAAGTCAAGAGGATGATAT 1940
 QY 108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValaValaSerTyrValaAsn 127
 DB 1941 TTGGAAGGACCAAGACCTCCCTTAAAGACATCAGTGAAGTCAAGTTCAGTAAAT 2000
 QY 128 ProGIyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
 DB 2001 CCTGGAAGGTGACAAAGCTGATGACCGTGGCTACTCATCTTGTCTGATAGTACAG 2060
 QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGluLysAsnProHisCysValSerLeu 167
 DB 2061 CACTGGAATCTCTTTTAAAGTATAGTAAACAGAAAGATCCACCTGCTGCTCCTT 2120
 QY 168 TyrLysLeuSerSerProGIyAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
 DB 2121 TACAAAGTATCAAGTCTCTGAAGTGAACCACTTGCAAAACAAAGAAATTTGGGCAACC 2180

```

QY      188  ILeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db      2181 ATTGGATTGACGACGAGTCTCTTCCTGACTATACCTCCAGAAATTTTCCTTTTGAA 2240
QY      208  SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db      2241 AGTACTACTGAGATTTCATTGTRTGGAAAGCTCTTACAAAGCTCATATCTACAGCTTGA 2300
QY      228  LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 247
Db      2301 AAGAAATATCCTACTGCTGCTGCTCATATATGGTGGTCCAGGTGAGTTGGGATTAAT 2360
QY      248  ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db      2361 CGGTTTAAAGAGATCAAGATTTCCTCCCTGAATACCTTACCTCTAGCTTATGTGGTT 2420
QY      268  ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyValPheLys 287
Db      2421 GTAGGATAGACACAGGGGATCTCTCACCGAGGCTTAAATTGAAAGCGCTTTTAA 2480
QY      288  TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db      2481 TATTAATG----- 2489
QY      308  ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyGlyTyr 327
Db      2489 ----- 2489
QY      328  LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db      2490 -----GTTGCTATTGCTGGGGCC 2507
QY      348  ProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
Db      2508 CCAGTCACCTCTGTGATCTTCTATGATACAGATACAG 2546

RESULT 13
BD157001      2161 bp      DNA      linear      PAT 17-JAN-2003
LOCUS      BD157001
DEFINITION      Primer for synthesizing full-length cDNA and use thereof.
ACCESSION      BD157001
VERSION      BD157001.1 GI:27862759
KEYWORDS      JP 2002191363-A/11844.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 2161)
AUTHORS      Ota,T., Isegai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
TITLE      Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
JOURNAL      Primer for synthesizing full-length cDNA and use thereof
PUBLISHED      Patent: JP 2002191363-A 11844 09-JUL-2002;
COMMENT      HELIX RESEARCH INSTITUTE
OS      Homo sapiens (human)
PN      JP 2002191363-A/11844
PD      09-JUL-2002
PF      28-JUL-2000 JP 2000280990
PI      TOSHIO OTA,TAKAO ISEGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KOJI
PI      SAITO,
PI      JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI      KEIICHI NAGAI,TETSUJI OTSUKI
PC      C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/
PC      10,
PC      C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
PC      Primer for synthesizing full-length cDNA and use thereof FH Key
FEATURES
source      Location/Qualifiers
FT      CDS      location(16)..(1911).
              1..2161
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"

```

```

BASE COUNT      662 a      429 c      452 g      618 t
ORIGIN
Alignment Scores:
Pred. No.:      1 216-148      Length:      2161
Score:      1528.00      Matches:      308
Percent Similarity:      57.79%      Conservative:      0
Best Local Similarity:      57.79%      Mismatches:      1
Query Match:      79.05%      Indels:      224
DB:      6      Gaps:      2
US-10-070-464-7 (1-360) x BD157001 (1-2161)
QY      1  GIUAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArg 20
Db      148  GAAAGAGATGCCAGATCAGCTGAGTCCGTAACCTTTGTTCTCCAAAGAAATTTGATAGA 207
QY      21  TyrSerGlyTyrTyrPyrCysProLysAlaGluThrThrProSerGlyGlyLysIleu 40
Db      208  TATTCGGCTATGGTGGTGGTCCAAAGCTGAAACAACTCCAGTGGGTAAATTTCTT 267
QY      41  ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
Db      268  AGAATTTCTATGAAAGAAATATGATGATCTGAGTGAATTAATTTATGATTTACATCCCT 327
QY      61  MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db      328  ATGTGGAAACAAGAGGGGACATTCATCCGTTATCTTAAACAGGTACAGCAAAATCCT 387
QY      81  LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle 98
Db      388  AAAGTCATTTTAAAGATGCAGAAATAATGATGATGTGTAAGAGAGATCATAGATGTC 447
QY      98  ----- 98
Db      448  ATAGATTAAAGAACTAATTCACCTTTTGAGATTCTATTGAAAGAGTTGAATATATTGCC 507
QY      98  ----- 98
Db      508  AGAGCTGATGAGACTCTGAGGAAATAATATGCTGTGTCATCTACTAGTGCCTCCAG 567
QY      98  ----- 98
Db      568  ACTGCCCTACAGATAGTGTGATCTCACCTGATTAATTAATCCAGTAGAAGATGATGTT 627
QY      98  ----- 98
Db      628  ATGGAAGGCAAGACTCATTTAGTCAGTGCCTGATTTGTGACGCCACTAATTAATCTAT 687
QY      98  ----- 98
Db      688  GAAAGAAACAAGACATCTGATTAATAATCATGACATCTTCATGTTTTCGCCAAGT 747
QY      98  ----- 98
Db      748  CAGGAAGAGAAATGAGTTATTTTGTGCTGTAATGCAAAAAGATTTCCGTCAATTA 807
QY      98  ----- 98
Db      808  TACAAATTAATCATCTATTTTTAAAGAAAGCAATATTAACGATCCAGTGGGTGCTCCT 867
QY      98  ----- 98
Db      868  GCTCCAGATGATTTCAAGTGTCTATCAAGAGAGATAGCAATTAACAGTGGTGAATGG 927
QY      99  -----ValAspGluValArgArgLeuValTyr 107
Db      928  GAAATTCCTGGCCGCGCATGATCTAATATCCAAAGTTGATGAGATGAGAGGCTGTGATAT 987
QY      108  PheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSerTyrValAsn 127
Db      988  TTTGAAGGACCAAGAAAGATCTCCCTTTAAGCATCACCTGTACTGACTGATTAAGTAAT 1047

```



```

QY 98 ----- 98
Db 508 AGAGCTGGATGACTCCTGAGGAAAAATATGCTTGGTCATCTACTAGATCGCTCCAG 567
QY 98 ----- 98
Db 568 ACTGGCTACAGATAGTGTGATCTCAGCTGAATATTTATCCAGTAGAAGATGATGTT 627
QY 98 ----- 98
Db 628 ATGAAAAGGAGAGACTCATTTAGTCAGTCTGATTTCTGTGACGCCACTAATATATAT 687
QY 98 ----- 98
Db 688 GAAAGAAACAACAGACATCTGGATPAATATCCATGACATCTTCAATGTTTTCCCAAAGT 747
QY 98 ----- 98
Db 748 CACGAAGAGAAATGAGTTATTTTTCCTCTGAATGCAAAAAGGTTCCGTCATTTA 807
QY 98 ----- 98
Db 808 TACAAATATACATCTATTTTAAAGAAAGCAATATPAAAGATCCAGTGGGGCTGCCT 867
QY 98 ----- 98
Db 868 GCTCCAGTGATTTCAAGTGTCTTATCAAAAGAGATGCAATTAACAGTGGTGAATGG 927
QY 99 ----- ValAspGluValArgArgLeuValTyr 107
Db 928 GAAGTCTTGGCCGGCATGATCTAATATCAAGTTGATGAAATCCAGAGGCTGGTATAT 987
QY 108 PheGluGlyThrIysAspSerProLeuGluHisIleuTyrValIleSerTyrValAsn 127
Db 988 TTTGAAGGACCAAAAGACTCCCTTAGAGCATCACCTGACAGTACAGTACGTAAT 1047
QY 128 ProGluValIleThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerArg 147
Db 1048 CCTGGAGAGGTGACAAAGGCTGACTGACCTGACCTGACATCTTGTGTCACAGCTAG 1107
QY 148 HisCysAspPhePheIleSerIysTyrSerAsnGluIysAsnProHisCysValSerLeu 167
Db 1108 CACTGTGACTTCTTTAATAGTATAGTAAACAGAAAGATCCACACTGTGTCTCTT 1167
QY 168 TyrIysLeuSerSerProGluAspAspProThrCysIysThrIysGluPheTyrAlaThr 187
Db 1168 TACAACTATCAAGTCTGAGATGACCCAACTTGCAAAACAAAGAAATTTTGCGCACCC 1227
QY 188 IleLeuAspSerIleGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db 1228 ATTTTGATTTCAAGAGGTCTCTCTCTGACTATATCTCTCCAGAAATTTTCTTTTGA 1287
QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIysProHisAspLeuGlnProGly 227
Db 1288 AGTACTACTGATTTACATTTGATGGAGATGCTCTACAGGCTCATATCTACAGCCTGA 1347
QY 228 TyrIysTyrProThrValIlePheIleTyrGlyGlyProGluValGlnLeuValAsnAsn 247
Db 1348 AAGAAATATCCCACTGTGTCTGATATATGTGGTCTCAGGTGACAGTGTGATTAAT 1407
QY 248 ArgPheIysGlyValIysTyrPheArgLeuAsnThrLeuIleSerLeuGlyTyrValIle 267
Db 1408 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACGCTCTTACGTTATGTGGTT 1467
QY 268 ValIleValIleAspAsnArgIysSerCysHisArgGlyLeuIysPheGluGlyAlaPheIys 287
Db 1468 GTAAGCATAGACAACAGGGGATCTCTGACCGAGGGCTTAAATTTGAAGCGGCTTTAA 1527
QY 288 TyrIysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db 1528 TATTAATATG----- 1536
QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 327

```

```

Db 1536 ----- 1536
QY 328 LeuSerIleuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db 1537 ----- GTTCTATTGCTGGGCC 1554
QY 348 ProValThrIleuTyrIlePheTyrAspThrGlyTyrThr 360
Db 1585 CCAGTCACTCTGTGATCTTCTATGATATACGATATACG 1593

```

```

RESULT 15
AX338497 2510 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 16 from Patent WO0183775.
DEFINITION AX338497
ACCESSION AX338497
VERSION 1
KEYWORDS AX338497.1 GI:18128919
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE
1 Deleane, A.M., Lal, P., Hafalia, A., Patterson, C., Walla, N.K., Kearney, L., Tribouley, C.M., Khan, F.A., Yao, M.G., Baughn, M.R., Azimzai, Y., Elliott, V.S., Nguyen, D.B., Gandhi, A.R., Yang, J., Hernandez, R., Policky, J.L., Lu, D.A., Reddy, R. and Tang, Y.T.
Proteases
Patent: WO 0183775-A 16 08-NOV-2001;
Incyte Genomics, Inc. (US)
Location/Qualifiers
FEATURES
SOURCE
1. 2510
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 376067CB1"
BASE COUNT 777 a 494 c 527 g 712 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 6,12e-148 Length: 2510
Score: 1522.00 Matches: 309
Percent Similarity: 58.08% Conservative: 0
Best Local Similarity: 58.08% Mismatches: 223
Query Match: 78.74% Indels: 2
DB: 6 Gaps: 2

```

US-10-070-464-7 (1-360) x AX338497 (1-2510)

```

QY 1 GluIuAspAlaArgSerAlaGlyValAlaThrPheValIleuGlnGluIupheAspArg 20
Db 609 GAAAGAGATGCCAGATCACTGAGTGGTACCTTTGTTCTCCAAAGAAATTTGATAGA 668
QY 21 TyrSerGlyTyrTyrProCysProIysAlaGluThrThrProSerGlyGlyIysIleu 40
Db 669 TATTCGGCTATTTGGTGGTGTCCAAAGCTGAAACAACTCCAGTGTGTAAATTCCT 728
QY 41 ArgIleLeuTyrGluIuAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
Db 729 AGAATCTATATGAGAAATATGATGATCTGAGTGGAAATTTATCATGTTACATCCCT 788
QY 61 MetLeuGluThrArg-ArgAlaAspSerPheArgTyrProIysThrGlyThrAlaAsnPr 80
Db 789 ATGTGGAACAAGGAGGAGGAGATTCATTCGTTATCTTAAACAGGTACACCAATCC 848
QY 80 GlyValThrPheIysMetSerGlnIleMetIleAspAlaGluGlyArgIleIle----- 98
Db 849 TAAAGTCACTTTAAGATGTCAGAATTAATGATGATGTGGAAGAGATCATATAGTGT 908
QY 98 ----- 98
Db 909 CATAGATAAGAACTAATTCAACCTTTTGAGATTCTATTGTAAGAGATTGAATATATTC 968

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:14:56 ; Search time 264.512 Seconds

(without alignments)
3673.932 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1933

Sequence: 1 BEDARAGVATFVLOEPRD.....RNAAGAPVTIMFTGTGT 360

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODER=frame+pzl.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US10070464/runat.15102003.113552.24799/bpp.query.fasta_1.2652
-DB=N Geneseq 15jun03 -QFMT=fastap -SUFFIX=mg -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=humand0.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10070464@cgn2.1_0@runat.15102003.113552.24799 -NCPUE=1 -ICPU=3
-NO_MAP -LAEGEODERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: N Geneseq 15jun03: *
2: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1980.DAT: *
3: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1981.DAT: *
4: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1982.DAT: *
5: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1983.DAT: *
6: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1984.DAT: *
7: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1985.DAT: *
8: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1986.DAT: *
9: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1987.DAT: *
10: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1988.DAT: *
11: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1989.DAT: *
12: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1990.DAT: *
13: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1991.DAT: *
14: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1992.DAT: *
15: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1993.DAT: *
16: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1994.DAT: *
17: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1995.DAT: *
18: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1996.DAT: *
19: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1997.DAT: *
20: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1998.DAT: *
21: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA1999.DAT: *
22: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT: *
23: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT: *
24: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA2001B.DAT: *
25: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA2002.DAT: *
26: /SIDSI/gcgdata/geneseg/geneseqn-emb1/NA2003.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1933	100.0	1083	22	Nucleotide sequenc
2	1880	97.3	4309	24	CDNA encoding huma
3	1836.5	95.0	2643	24	Coding sequence of
4	1836.5	95.0	2649	25	CDNA encoding huma
5	1836.5	95.0	2671	24	CDNA encoding huma
6	1836.5	95.0	3106	24	CDNA encoding huma
7	1836.5	95.0	3120	22	Human protease PRT
8	1836.5	95.0	3120	24	Nucleotide sequenc
9	1836.5	95.0	3143	24	Human dipeptidyl p
10	1816.5	94.0	4829	24	CDNA encoding 2195
11	1528	79.0	2151	22	CDNA encoding huma
12	1522	78.7	2510	22	Human CDNA sequenc
13	1522	78.7	2510	24	Human CDNA sequenc
14	1521.5	78.7	4676	24	CDNA encoding huma
15	1429	73.9	1821	24	CDNA encoding huma
16	1254.5	64.9	2668	24	Dipeptidyl peptida
17	1254.5	64.9	2842	24	Novel human coding
18	1229	63.6	4523	24	Novel human coding
19	1220.5	63.1	2495	24	CDNA encoding huma
20	1220.5	63.1	2617	24	Human dipeptidyl p
21	1220.5	63.1	3024	24	CDNA encoding huma
22	1220.5	63.1	3716	24	Human dipeptidyl p
23	1220.5	63.1	4219	24	Human PMW encodin
24	1220.5	63.1	4302	24	CDNA encoding huma
25	1204.5	62.3	2751	24	CDNA encoding huma
26	1204.5	62.3	3287	24	Murine dipeptidyl
27	1201	62.1	4180	24	Alternative versio
28	1183.5	61.2	4263	24	CDNA encoding huma
29	1183.5	61.2	3282	22	CDNA encoding huma
30	1169.5	60.5	2801	22	Human polynucleoti
31	1169.5	60.5	2982	22	Human polynucleoti
32	1132.5	58.6	4076	24	Human polynucleoti
33	1132.5	58.6	4159	24	CDNA encoding huma
34	1113	57.6	4037	24	CDNA encoding huma
35	1113	57.6	4120	24	CDNA encoding huma
36	1086.5	56.2	1197	22	CDNA encoding huma
37	984	50.9	2952	24	Nucleotide sequenc
38	980	50.7	3047	24	DNA encoding huma
39	944	48.8	2027	21	Human ORFX ORF2692
40	860.5	44.5	1669	22	Nucleotide sequenc
41	722	37.4	3713	23	Drosophila melanog
42	722	37.4	3783	23	Drosophila melanog
43	654.5	33.9	2461	21	Human ORFX ORF1390
44	558	28.9	6225	23	Drosophila melanog
45	558	28.9	6228	23	Drosophila melanog

ALIGNMENTS

RESULT 1
ID AAC85697
AAC85697 standard; cDNA; 1083 BP.
ID AAC85697;
29-JUN-2001 (first entry)
Nucleotide sequence encoding human DPP8 244Glu-341Ile+515Val-776Thr.
AC Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
DE dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
XX growth hormone deficiency; glucose level; mucosal regeneration;
KW non-insulin dependent diabetes mellitus; glucose intolerance;
KW immunosuppression; ss.
OS Homo sapiens.

XX	FH	Key	Location/Qualifiers
XX	FT	CDS	2..1081
XX	FT	/product= "Human DPP8	P244Glu-341Ile+515Val-776Thr"
XX	FT	/tag= a	
XX	FT	/notes= "No stop codon given"	
XX	FT		
XX	PN	WO200119866-A1.	
XX	PD	22-MAR-2001.	
XX	PF	11-SEP-2000; 2000MO-AU01085.	
XX	PR	10-SEP-1999; 99AU-0002762.	
XX	PR	18-FEB-2000; 2000AU-0005709.	
XX	PA	(UNSY) UNIV SYDNEY.	
XX	PI	Abbott CA, Gorell MD;	
XX	XX	WPI: 2001-281520/29.	
XX	DR	P-PESDB; AAB47190.	
XX	PT	New human dipeptidyl aminopeptidase (DPP8) useful for cleaving	
XX	PT	substrates, identifying inhibitors of DPP8 catalytic activity which	
XX	PS	have therapeutic uses, and for detecting activated T cells -	
XX	PS	Claim 21; Page 75-76; 78pp; English.	
CC	XX	The sequences given in AAC85695-97 encode fragments of human	
CC	CC	dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for	
CC	CC	H-Ala-Pro-pNA, H-Gly-Pro-pNA and H-Arg-Pro-pNA. Therefore, it is a	
CC	CC	prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable	
CC	CC	of hydrolysing the peptide bond C-terminal to Pro in each of these	
CC	CC	compounds. DPP8 is homologous with human DDPV. DPP8 is useful for	
CC	CC	cleaving a substrate, and for detecting an activated T cell which for	
CC	CC	involves measuring the level of DPP8 gene expression in a T cell. The	
CC	CC	level of DPP8 expression is detected by detecting the amount of DPP8	
CC	CC	RNA in the cell. It is also useful for identifying a molecule capable	
CC	CC	of inhibiting the cleavage of the substrate by DPP8. Molecules	
CC	CC	identified as inhibiting DPP8 catalytic activity may be useful for	
CC	CC	treating diarrhoea, growth hormone deficiency, lowering glucose levels	
CC	CC	in non-insulin dependent diabetes mellitus and other disorders	
CC	CC	involving glucose intolerance, enhancing mucosal regeneration and	
CC	XX	as immunosuppressants.	
SQ	XX	Sequence 1083 BP; 312 A; 213 C; 247 G; 311 T; 0 other;	
<hr/>			
Alignment Scores:			
Pred. No.:		1,46e-216	Length: 1083
Score:		1933.00	Matches: 360
Percent Similarity:		100.00%	Conservative: 0
Best Local Similarity:		100.00%	Mismatches: 0
Query Match:		100.00%	Indels: 0
DB:		22	Gaps: 0
<hr/>			
US-10-070-464-7 (1-360) x AAC85697 (1-1083)			
OY	1	GjngJuaSPAlAargSerAlagIyValAlaThrPheValIengJngJngIuePaaspArg	20
Dd	2	GAAAGAAGATGCCAGATCAGCTGGAGTCGCACCTTGTCTCCAGGAAGATTGATAGA	61
OY	21	TyrserGIyTyTrphrPCysProIysAlaGluThrThrProserGIyGLyISlleu	40
Dd	62	TATCTGGCTATTGGTGtGTCcAAAAGCGTGAACAACATCCAGTGtGGTAAAAATCTT	121
OY	41	ArgIleLeurTyGIuJngJuaSnAspGiuSerGIuValGlunIleIehISValThnSerPro	60
Dd	122	AGAATTCTATATGAAAGAAATGATGATCTGAAGTGGAATTATTAATCATCCCT	181
OY	61	MetLeuGIuThrArgArgAlaaspSerPheArgTyProIySthrGIyThrAlaasnPro	80
Dd	182	ATgTTGGAAACAAGGGGCGAGTTATTCCTGTTATCTTAAACAGGTACAGCAAACTCT	241

QY	81	LysValThrPheIysMetSerGlnIleMetIleAspAlaGluGlyArgGlieIleValAsp	100
Db	242	AAAGTCACCTTTTAAGATGTCAGAAATATATGATGATCTGAAGGAAGGATCATAGTTGAT	301
QY	101	GluValAlaGlyLeuValTyrPheGluGlyThrIleAspSerProLeuGlnHisIleu	120
Db	302	GAAGTCAGAGGCTGGTATATTGTAAGGACCAAGACCTCCCTTTAGAGCATCACCTG	361
QY	121	TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer	140
Db	362	TACGAGTCACTTACGTAAATCTCGAGAGGTGCAAGGCTGACTGACCGGTACTCA	421
QY	141	HisSerCysCysIleSerGlnHisCysAspPheIleSerIlyTyrSerAsnGlnIys	160
Db	422	CATCTGCTCCATCAGTCAGACACCTGAGACTCTCTTTTAAGTAAGTATAGTAAACAAG	481
QY	161	AsnProHisCysValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysIys	180
Db	482	AATCCACACTGTGTGTCCTTTACAGCATACAGTCTTGAAAGTGCACCACTTGCAA	541
QY	181	ThrIysGlnPhePheIleThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro	200
Db	542	ACAAAGGATTTTGGGCGACCATTTTGATTCACACAGTCCCTCTCCGACTATATACCT	601
QY	201	ProGlnIlePheSerPheGlnSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIys	220
Db	602	CCAGAAATTTCTCTTTGAAAGTACACAGTATTCATTGTATGGAGTCTCTTACAG	661
QY	221	ProHisAspLeuGlnProGlyIleIysTyrProThrValLeuPheIleTyrGlyIlePro	240
Db	662	CCTCATGATCTACAGCCTCGAAGAAATATCTTCTGTGCTGTTCAATATATGTGTGCT	721
QY	241	GlnValGlnLeuValAsnAsnArgPheIysGlyValIysTyrPheArgLeuAsnThrLeu	260
Db	722	CAGGTCAGTGGTGAATTAATCGGTTTAAAGAGTCAAGATTTCCGCTTAATACCTA	781
QY	261	AlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeu	280
Db	782	GCCCTCTAGGTATATGCGTTGTATGATATGATACAAACAGGGGATCTGTACACGAGGCTT	841
QY	281	LysPheGluGlyAlaPheIysTyrIleMetGlyGlnIleGluIleAspAspGlnValGlu	300
Db	842	AAATTTGAAGCGCCTTTAAATATAAATGGGTCAAAATGAAATTGACGATCAGTGGAA	901
QY	301	GlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHis	320
Db	902	GGACTCCCAATATCTAGCTTCGATATGATTTCAATGACTTAGCTGCGTGGCATCAC	961
QY	321	GlyTyrPheSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePhe	340
Db	962	GCGGTGTCATGAGAGATACCTCTCCCTGATGCAATTAAGCAGAGTCCGATATCTTC	1021
QY	341	ArgValAlaIleAlaGlyAlaProValThrLeuThrPheTyrAspThrGlyTyrThr	360
Db	1022	AGGGTGTATTTGCTGGGGCCCCGACGTCATCTGTGCAATCTTCTATGATACGATACAG	1081
RESULT 2			
ABK83328			
ID	ABK83328 standard; cDNA; 4309 BP.		
XX	ABK83328;		
AC			
XX	12-AUG-2002 (first entry)		
DT			
XX	cDNA encoding human DPRP-1 splice variant #4.		
DE			
XX	Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP; DPPYV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain; diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke; heart failure; hypertension; urinary retention; osteoporosis; cancer; ulcer; allergy; cancer; psychotic disorder; neurological disorder; dyskinesia; reproductive disorder; inflammatory disorder;		

KW metabolic disorder; gene; ss.
 XX Homo sapiens.
 OS
 XX WO200231134-A2.
 PN
 XX 18-APR-2002.
 PD
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI: 2002-444178/47.
 DR P-PSDB: ABG61597.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 PS Disclosure: Page 67-68; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyslexias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABR83322-ABR83343 encode human DPPR proteins.
 XX
 SQ Sequence 4309 BP; 1304 A; 790 C; 907 G; 1308 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 1.95e-209 Length: 4309
 Score: 1880.00 Matches: 358
 Percent Similarity: 98.90% Conservative: 0
 Best Local Similarity: 98.90% Mismatches: 2
 Query Match: 97.26% Indels: 3
 DB: 24 Gaps: 0
 US-10-070-464-7 (1-360) x ABR83328 (1-4309)
 QY 1 GIUGIUAAPALAAAGSeraIAGIYValAlaThrPheValIleuGInGIuIuPheAspArg 20
 DB 943 GAAGAGATGCCAGATCGAGTCGCTACCTTGTCTCCAGAGAAATTTATAGA 1002
 QY 21 TYSerGIYTYRTIPTCysProIysAlaGIuIuThrProSerGIYGIYLSIleu 40
 DB 1003 TATCTGGCTATGTGGTGTGCCAAAGCTGMAAACACTCCAGGCGTAAATTCCT 1062
 QY 41 ArgIleuTYRGIUGIUAAsnAspGIuSerGIUValIleuIleHisValaThrSerPro 60
 DB 1063 AGAATCTTATATGAAGAAATGATGATCTGAGGGAATATATTCAGATTACATCCCT 1122
 QY 61 MetIeuGIuIuThrArgAlaAspSerPheArgTYRProIysThrGIYThraIAsnPro 80
 DB 1123 ATGTTGAAACAGAGAGGAGATTCATTCGTTATCTTAACACGATACGAAATCTT 1182
 QY 81 LysValIThrPheIysMetSerGIuIleuMetIleAspAlaGIUGIYArgIleIleValAsp 100
 DB 1183 AAAGTCACTTTTAAGATGTCAAGAAATATGATGATGCTGAAGAGAAATCCAA-GTTGAT 1241

QY 101 GIUValAGArgLeuValITyrPheGIUGIYThrIYAspSerProIeuGIuHisIleu 120
 DB 1242 GAAGTCAGAAAGGCTGTATATTTGAAGGCACCAAGACTCCCTTTAGAGCATCACCTG 1301
 QY 121 TYRValIAserTYRValAsnProGIUGIuValIThrArgLeuThrAspArgIYTYRser 140
 DB 1302 TACGTAGTCAGTTAGCTAAATCTTGAAGAGGTGAACAGGCTGACGAGCTGACTCA 1361
 QY 141 HisSerCYSsIleSerGlnHisCysAspPhePheIleSerIYsSerAsnGlnIys 160
 DB 1382 CATTCTTGCTGATCAGTCAGACCTGTGACTCTTTATATAGATAGATATGTAACCGAAG 1421
 QY 161 AsnProHisCysValSerIeuTYRlysIeuSerSerProGIUAspAspProThrCysIys 180
 DB 1422 AATCCACACTGTGTCTCCCTTTACAAAGTATCAAGTCTCGAAGATGACCAACTGGCAA 1481
 QY 181 ThrIysGIuPheThrAlaThrIleIeuAspSerAlaGIYProIeuProAspTYRThrPro 200
 DB 1482 ACAAAAGAAATTTGGGCCACCAATTTGGATTAGAGAGGTCTCTTCTGACTATACTCT 1541
 QY 201 ProGIuIlePheSerPheGIuSerThrThrGIYPheThrLeuTYRGIYMetLeuTYRlys 220
 DB 1542 CCAGAAATTTCTCTTTGAAAGTACTAGATTTACATTTGATGGATGCTCTACAG 1601
 QY 221 ProHisAspLeuGlnProGIYLYSLysTYRProThrValIeuPheIleTYRGIY--P 240
 DB 1602 CCTCATGATCTACACCTGGAAAGAAATCTACTGTGCTCTTATATATGATGTGCTTC 1661
 QY 240 roGIuValGIuIleuValAsnAsnArgPheIysGIYValIysTYRPhaArgLeuAsnThrL 260
 DB 1662 CTCAGGTGAGTTGGTGTATATTCGTTTAAAGAGTCAAGATATTCCTGTAATACCC 1721
 QY 260 euAlaSerIeuGIYTYRValValValIleAspAsnArgIYserCysHisArgGIYL 280
 DB 1722 TAGCTCTCTAGTATAGGTATGATGATAGATAGACACAGGGAGATCTCTACCGAGGCC 1781
 QY 280 euIysPheGIUGIYAlaPheIysTYRlysMetGIYGlnIleGIuIleAspAspGIuValG 300
 DB 1782 TTAAATTTGAAGGCCCTTTAAATATTAAGATGGTCAATATGAATTCAGATCAGTGG 1841
 QY 300 IuGIYLeuGIuTYRleuAlaSerArgTYRAspPheIleAspIeuAspArgValIleH 320
 DB 1842 AAGGACTCAATATTTAGCTTCTCGATATGATTTATGACTTATGATCTGTGGGACTCC 1901
 QY 320 IsgIYTYRserTYRGIYGIYTYRleuSerIeuMetAlaIeuMetGlnArgSerAspIleP 340
 DB 1902 ACGGCTGTCTTATGAGAGATACCTCTCCCTGATGCGATTTATGACAGAGTCAATATCT 1961
 QY 340 heArgValAlaIleAlaGIYAlaProValIThrLeuTYRlePheTYRAspThrGIYTYT 360
 DB 1962 TCAGGGTTGCTATGTGCTGGGCCCCAGTCACCTGTGTGATCTTCTATGATACAGATACA 2021
 QY 360 hr 360
 DB 2022 CG 2023
 RESULT 3
 AAH99935
 ID AAH99935 standard; cDNA, 2643 BP.
 XX
 AC AAH99935;
 XX
 DT 12-APR-2002 (first entry)
 XX
 DE Coding sequence of 21953 human prolyl oligopeptidase.
 XX
 KW 21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
 KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;
 KW neuronal disorder; vascular disorder; prostate disorder; cyrostatic;
 KW antidiabetic; antianthratic; antiaslthmatic; antinflmmatory;
 KW diabetes mellitus; arthritic; multiple sclerosis; asthma;
 KW Grave's disease; neuronal disorder; demyelinating disease; ss.
 XX

OS Homo sapiens.
XX
XX WO200179473-A2.
XX
XX 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-US40483.
XX
XX 18-APR-2000; 2000US-197508P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Meyers RA, Williamson M;
XX
XX WPI, 2002-034353/04.
XX
XX P-PSDB; AAG78415.
XX
XX New polypeptides 21953, member of human prolyl oligopeptidase family,
PT useful as diagnostic targets and therapeutic agents for controlling
PT cancer, lymphoma and leukemia -
XX
XX
XX Claim 2; Page: 121pp; English.
XX
XX This invention relates to an isolated 21953 human prolyl
XX oligopeptidase, which is cytosolic, antidiabetic, antiarthritic,
XX neuroprotective, antihypertensive, dermatological, antipsoriatic,
XX antiaesthetic, ophthalmological, antiinflammatory, nociceptive,
XX antiparkinsonian, anticonvulsant, gynecological, vasotropic,
XX antidiabetic, antidiabetic, antidiabetic, anorectic and
XX antidiabetic in its action. Uses include gene therapy, expression or
XX activity of 21953 protein modulator, it is useful for identifying a
XX compound which binds to it and can be used in preventing, treating
XX or detecting a cellular proliferative or differentiative disorder.
XX The 21953 molecules can act as novel diagnostic targets and therapeutic
XX agents for controlling disorders associated with the aberrant activity
XX or degradation of peptide hormones e.g., disorders associated with cell
XX differentiation and proliferation such as cancer, immune function,
XX reproductive, neurological and cardiovascular function. The 21953
XX molecules are thus useful for treating and preventing cellular
XX proliferative and differentiative disorders, haematopoietic neoplastic
XX disorders, immune disorders such as autoimmune diseases, diabetes
XX mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
XX neuronal disorders, demyelinating diseases, vascular disorders and
XX metabolism or pain disorders. This sequence represents the cDNA
XX encoding sequence of 21953 human prolyl oligopeptidase. This
XX sequence represents the coding sequence of 21953, being the sequence
XX in between the start and the stop codon of the sequence represented in
XX AAH99934.
XX
XX Sequence 2643 BP; 800 A; 514 C; 585 G; 744 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1 19e-204 Length: 2643
Score: 1836.50 Matches: 360
Percent Similarity: 67.54 Conservative: 0
Best Local Similarity: 67.54 Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 24 Gaps: 1
US-10-070-464-7 (1-360) x AAH99935 (1-2643)
QY 1 GIUGIUSPALAARYSERIALGIYVALAATHRPHEVALLEUGINGLUNIPHEASPAR 20
DB 727 GAAGAAGATGCCAGATCACTGAGAGTCCTACCTTTGTTCTCCAGAAAGATTTGATAGA 786
QY 21 TYRSEGIYTYRTTIPCYSPROLYSALGIUTNRTHPROSEGIYGLIYLSILEU 40
DB 787 TATTCGCTATTGGTGTCGTCACAAAGCTGAACCACTCCAGAGTGCTGTAATAATCTT 846
QY 41 ARGILEUITYRGILUGIUSNAPSGIUSERGIUVALGIUILEIETHISVALTHRSEPRO 60
DB 847 AGAATTCTATATGAGAAATATGATGATGTGAGGTGAGAAATATTCATTTTCACTCCCT 906

QY 61 MELLEUGLUTHRARGARGHIAASPSERPHARGTYRPROLYETHRGILYHRAIAENPRO 80
DB 907 ATGTGGAAACAAAGAGGGCAGATTATCCGTATCTTAAACAGGTACAGAAATCCCT 966
QY 81 LYSVALTHRPHELYMERSEGLIILEWETILEASPAJAGIUGIYARGILEILE 98
DB 967 AAGTCACCTTTTAAGTATGTCAGAAATATGATTTGATGCTGAAAGAGATCATAGTGC 1026
QY 98 ----- 98
DB 1027 ATAGATAAGAACTAATTCACCTTTGAGATTCTATTGGAAGAGTGAATATATTGCC 1086
QY 98 ----- 98
DB 1087 AAGCTGATGAGATCTCGAGGAAATATGCTTGTCATCTACTAGATGCTCCAG 1146
QY 98 ----- 98
DB 1147 ACTGCGCTGCAATAGTGTGATTCACCTGAATTTATTCACAGATAGATGATT 1206
QY 98 ----- 98
DB 1207 ATGAAAGGCAAGACTCATGATGATGAGTCTGATTCGTGACGCCACTAATATCTAT 1266
QY 98 ----- 98
DB 1267 GAAGAAACACAGACATCTGATTAATATCATGACATCTTCATGTTTCCCCAAAGT 1326
QY 98 ----- 98
DB 1327 CACGAAAGAAATGAGTTATTTTGGCTGTGAATGCAAAACAGTTCCGCTATTAA 1386
QY 98 ----- 98
DB 1387 TACAAAATTACATCTATTAAAGAAAGCAATATTAACGATCCAGTGTGCTGCT 1446
QY 98 ----- 98
DB 1447 GCTCCAAAGTATTCAAGTCTCTTCAAAAGAGATGCAATTAACAGTGTGATGG 1506
QY 99 ----- 98
DB 1507 GAAGTCTCTGGCGGCGATGATCTAATATCCAAAGTTGATGAAGTGAAGGCTGATAT 1566
QY 108 PHEGLUGLTHLYSASPSERPROLEUGIUSHISLEUITYRVALYISERTYRVALAAN 127
DB 1567 TTTGAAGGCAACCAAGACCTCCCTTTAGACATCACTGTAGTGTGATTAAGTAAAT 1626
QY 128 PROGLYGLIUALTHRARGLEUTHRASPARGIYTYRSETHISERCYSIIESEGIN 147
DB 1627 CCTGAGAGGTATCAAGGCTGATGCTGACCTGCTGCTCATCTCTTCTGATCAGTCA 1686
QY 148 HISYASPPHEPHEILESERIYTYRSEANGILYASNPROIHCYSVALSERLEU 167
DB 1687 CACTGTGACTCTTTATTAAGTATAGTATGTAACCAAGAAATCCACAGTGTGCTCTT 1746
QY 168 TYRYSLEUSERSEPROGLIUSASPSERPROTHCYSLYTHLYGLIUNHETPAATMR 187
DB 1747 TACAAGCTATCAAGCTCGAAGATGACCAACTCCCAAAACAAAGAAATTTGGGCGACC 1806
QY 188 ILELEUSPSEALGIYPROLEUPROASPIYRTHPROPROGLIUILEPHESEPRGIN 207
DB 1807 ATTTGGATTCAGCAGGTCCTCTTCTGATCATCTCCCGCAAAATTTCTCTTTGAA 1866
QY 208 SERTHRGILYRPHETHEUITYRGLIYMETLEUITYRYSPROHISASPLEUNGPROGIY 227
DB 1867 AGTACTACTGATTTACATTTGATGAGATGCTCTCAAGCCCATGATCTACAGCTGGA 1926
QY 228 LYSLSYTYRPROTHVALLEUPHEIILETYRGIYGLIYPROGINVALGLEUVALJASNA 247
DB 1927 AAGAAATTCCTACTGTGTCTCATATATAGTGTCCTCAAGTGTGAGTGTGAAATAT 1986
QY 248 ARGPHELYSGIYVALYTYRPHARGLEUSNTHRLEUALSERLEUGIYTYRVALVAL 267


```
Db 1330 CACGAGAGAAATTGAGTTATTTTGGCTCTGATGCAAAACAGGTTCCGTCATTTA 1389
QY 98 ----- 98
Db 1390 TACAAAATTACATCTATTTTAAAGGAAACAAATATAACGATCCAGTGTGGCTGCT 1449
QY 98 ----- 98
Db 1450 GCTCCAGATGATTCAAGTTCCTTCAAGAAGAGATGACAAATTACAGTGGTGAATGG 1509
QY 99 -----ValAspGluValArgLeuValTyr 107
Db 1510 GAAGTCTTGGCCGCGCATGATCTAATATCCAGTTGATGAAGTGAAGGCTGGTATAT 1569
QY 108 PheGluGlyThrLeuAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
Db 1570 TTGGAAGGACCAAAAGACTCCCTTTAGAGCATCAGTCGTGAGTGCATGACGTAAT 1629
QY 128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysSerLeuSerGln 147
Db 1630 CCTGGAGAGGTGACAAAGGCTGACTGACCGTGGCTACTACATTTCTTGCTGACGCTAG 1689
QY 148 HisCysAspPhePheLeuSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db 1690 CACTGTGACTTCTTATATAGTATAGTATAGTAAACGAAGAATCCACACTGTGTGCTT 1749
QY 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr 187
Db 1750 TACAAGCTATCAAAAGCTCGAAACATACCCAACTTCCAAACAAAGAAATTTGGCCAC 1809
QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluHisPheSerPheGlu 207
Db 1810 ATTTGGATTCAAGGAGGCTCTTCTGACTATCTCTCCGAAATTTCTCTTTGAA 1869
QY 208 SerThrThrArgLysPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db 1870 AGTACTACTGAGATTACATTTGATGAGATGCTCTCAAGACCTCATGATCTACAGCTGGA 1929
QY 228 LysLysTyrProThrValLeuPheLeuTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
Db 1930 AAGAATATCTTACTAGTGTGCTTCAATATATGATGCTTCAAGTGCAGTGTGATATAT 1989
QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db 1990 CGGTTTAAAGAGTCAAGATTTCCGCTTGAATACCTTACCTCTAGGTTATGTGTT 2049
QY 268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyValaPheLys 287
Db 2050 GTAGGATAGACAAACAGGGGATCCGTCAACCGAGGGCTTAAATTTGAAGGCGCTTTAAA 2109
QY 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db 2110 TATATAAATGGTCAATATGAATATGACATCAGTGAAGGAGCTCCAAATATCTACCTCT 2169
QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyLysTyr 327
Db 2170 CGATATGATTTTCATTGACTTAGATCGTGTGGGCATCCAGCTGTGTCATAGAGAGATAC 2229
QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db 2230 CTCTCCCTGATGGCATATATGACAGATATCTTCAAGGCTGCTATGTCTGGGCGCC 2289
QY 348 ProValThrLeuTyrIlePheTyrAspThrArgLysTyrThr 360
Db 2290 CCAGTCACTCTGTGATCTTCTATATACAGATACAG 2328
RESULT 5
ABK83322
ID ABK83322 standard; cDNA; 2671 BP.
XX AC
XX ABK83322;
```

```
DT 12-AUG-2002 (first entry)
XX 12-AUG-2002 (first entry)
DE cDNA encoding human DPPIV related serine protease DPP-1.
XX
XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
XX DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
XX diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
XX heart failure; hypertension; urinary retention; osteoporosis; cancer;
XX ulcer; allergy; cancer; psychotic disorder; neurological disorder;
XX dyskinesia; reproductive disorder; inflammatory disorder;
XX metabolic disorder; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200239134-A2*
XX
XX 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US31874.
XX
XX 12-OCT-2000; 2000US-240117P.
XX
XX (FERR ) FERRING BV.
XX
XX Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
XX WPI. 2002-444178/47.
XX
XX P-PSDB; ABG61591.
XX
XX
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
XX the proteins; useful for treating e.g. fungal, bacterial, protozoan and
XX viral infections, cancers, allergies, neurological disorders, or pain
XX
XX
XX Claim 1; Page 53-54; 113pp; English.
XX
XX
XX The present invention relates to the isolation of novel human serine
XX proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
XX proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
XX and nucleic acids encoding them are useful for treating infections
XX such as fungal, bacterial, protozoan and viral infections, particularly
XX infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
XX cystitis, diabetes, precocious puberty, infertility, obesity, anorexia,
XX bulimia, Parkinson's disease, acute heart failure, hypotension,
XX hypertension, urinary retention, osteoporosis, angina pectoris,
XX stroke, ulcers, asthma, allergies, cancers, migraine, vomiting, or
XX psychotic and neurological disorders (e.g. anxiety, dementia, or
XX schizophrenia), and dyskinesias. These may also be used in discovering
XX therapeutic agents for the treatment of reproductive, inflammatory and
XX metabolic disorders. ABK83322-ABK83343 encode human DPPR proteins.
XX
XX
XX Sequence 2671 BP; 805 A; 524 C; 594 G; 748 T; 0 other;
XX
XX
XX Alignment Scores:
XX
XX Pred. No.: 1,21e-204 Length: 2671
XX Score: 1836.50 Matches: 360
XX Percent Similarity: 67.54% Conservative: 0
XX Best Local Similarity: 67.54% Mismatches: 0
XX Query Match: 95.01% Gaps: 173
XX DB: 24 Indels: 1
XX
XX US-10-070-464-7 (1-360) x ABK83322 (1-2671)
QY 1 GlnGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
Db 737 GAGGAAAGTGCAGATCAGCTGAGATCCCTTCTTGTTCGAAGAAATTTATATGA 796
QY 21 TyrSerGlyTyrTyrProCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
Db 797 TATTCGCTATTTGGTGTGTGCCAAAGCTGAAGAACATCCAGAGCTGTAAATTCCT 856
QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleHisValThrSerPro 60
```


Db 857 AGAATTCCTATATGAGAAAATGATGATCTGAGGTGGAATATTATTCATCTCCCT 916
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProIystrHrgIYThrAlaAsnPro 80
 Db 917 ATGTGGAACAAGAGAGGAGATTCATTCGGTATCTCTAAACAGGATGACAAATCCT 976
 QY 81 LysValIlePheIlePheMetSerGluIleMetIleAspAlaGluGlyValIle----- 98
 Db 977 AAAGTCACTTTTAAAGATGTGAGAAATATATGATGATGCTGAAGAGAGATCATGATGTC 1036
 QY 98 ----- 98
 Db 1037 ATAGATAGAGAACTAATTCACCTTTGAGATTCATTTTGAAGAGATTGAATATATGTC 1096
 QY 98 ----- 98
 Db 1097 AGAGCTGATGAGCTCCTGAGAGAAAATATGCTTGTCATCTCTAATAGATGCTCCAG 1156
 QY 98 ----- 98
 Db 1157 ACTGCGCTGAGATAGTGTGATCTCACCTGAATATTTATCCAGAGAGATGATGTT 1216
 QY 98 ----- 98
 Db 1217 ATGGAAGGAGAGACTCATGAGTCAAGTGCCTGATTCGTGACGCCACTAATATCTAT 1276
 QY 98 ----- 98
 Db 1277 GAAGAAACAACAGACATCTGATTAATATCCATGACATCTTTCANGTGTTTCCCAAAGT 1336
 QY 98 ----- 98
 Db 1337 CACGAGAGAGAAATGAGTTATTTTGGCTCTGATGCAAAACAGGTTTCGTCATTTA 1396
 QY 98 ----- 98
 Db 1397 TACAAAATTACATCTATTTTAAAGAAAGCAATATTAACGATCCAGTGGTGGCTGCT 1456
 QY 98 ----- 98
 Db 1457 GCTCAAGTATTCATAGTGTCTATCAAAAGAGAGATGCAATTAACAGTGTGATGAG 1516
 QY 99 -----ValAspGluValAlaArgLeuValTyr 107
 Db 1517 GAAGTCTTGCCGCGCATGATCTAATATCCAAAGTTATGATGAGAGAGCTGGTATAT 1576
 QY 108 PheGluGlyThrIleAspSerProLeuGluIleHisIleuTyrValIleSerTyrValAsn 127
 Db 1577 TTTGAGGCGACCAAGACTCCCTTTAGAGCATCACTGTACGATGATGATGAT 1636
 QY 128 ProGluGluValIleThrArgLeuThrAspArgGlyTyrSerHisSerCysAsiIleSerGln 147
 Db 1637 CCTGAGAGAGGTGACAAAGCTGACGACGCTGATCACTCTGTCGATCAGTCAG 1696
 QY 148 HisCysAspPhePheIleSerIleTyrSerHisIleAsnProHisCysValSerLeu 167
 Db 1697 CACTGATCACTCTTATATAGATATATAGATATACAGAAATCCACCTGTGTCTCTT 1756
 QY 168 TyrIleuSerSerProGluAspAspProThrCysIleTyrIleGluPheTyrAlaThr 187
 Db 1757 TACAGCTATCAATCTCTGAAGATGACCAACTTGCAAAACAAAGAAATTTGGGCCACC 1816
 QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207
 Db 1817 ATTTTGATTCAGAGAGTCTCTCTGACATATCTCTCCAGAAATTTTCTTTGAA 1876
 QY 208 SerThrThrGlyPheThrIleuTyrGlyMetLeuTyrIleProHisAspLeuGlnProGly 227
 Db 1877 AGTACTACTGATTTACATTTATATGAGATGCTCTCAAGCTCAATGATCAACCTCGGA 1936
 QY 228 LysIleTyrProThrValIleuPheIleTyrIleGlyProGluValIleGluValIleAsn 247
 Db 1937 AAGAAATATCTACTGTGCTGTTCATATATGATGCTCTCTCAAGGTCAGTTGGTGAATAT 1996

QY 248 ArgPheIleGlyValIleTyrPheArgLeuAsnThrIleuAlaSerLeuGlyTyrValIle 267
 Db 1997 CGATTTAAAGAGAGCAAGTATTTCCGCTTGAATACCTAGCCTCTTAGGTTATGTGTT 2056
 QY 268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuIlePheGluValIlePheIle 287
 Db 2057 GTAGTGTATGACCAACAGAGGATCTGTACCGAGAGGCTTAAATTTGAAGGGCCCTTAAA 2116
 QY 288 TyrIleMetGlyGlnIleGluIleAspAspGlnValIleGluIleGlnTyrLeuAlaSer 307
 Db 2117 TATTAATGGGTCAATATGAAATGACATGATGAGTGAAGAGCTCCATATCTAGCTTCT 2176
 QY 308 ArgTyrAspPheIleAspLeuAspArgValIleHisGlyTyrPheTyrGlyValIle 327
 Db 2177 CGATATGATTTTCAATGATCTTAGATCGTGGGACATCCAGCGCTGCTCTATGAGAGATAC 2236
 QY 328 LeuSerIleMetAlaLeuMetGlnArgSerAspIlePheArgValIleAlaGlyValIle 347
 Db 2237 CTCCTCCGATGGGATTAATGACAGAGTCAAGATATCTTCAAGGTTGCTATTTGCGGSC 2296
 QY 348 ProValThrLeuTyrPheTyrAspThrGlyTyrThr 360
 Db 2297 CCAATCACTCTGTGATCTTCTATGATACAGATACAG 2335
 RESULT 6
 ID ABR12892 standard; cDNA; 3106 BP.
 AC ABR12892;
 XX 09-APR-2002 (first entry)
 DT Human protease PRTS-9 cDNA sequence.
 DE Human protease PRTS-9 cDNA sequence.
 XX Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
 KM cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
 KM inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
 KM cell proliferative disorder; developmental disorder; epilepsy;
 KM Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
 KM reproductive disorder; endometriosis; ss.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 203..2851
 FT /*tag= a
 FT /product= "Human protease PRTS-9"
 FT
 XX MO200198468-A2.
 XX
 PD 27-DEC-2001.
 XX
 PE 13-JUN-2001; 2001WO-US19178.
 XX
 PR 16-JUN-2000; 2000US-212336P.
 PR 22-JUN-2000; 2000US-213955P.
 PR 29-JUN-2000; 2000US-215396P.
 PR 07-JUL-2000; 2000US-216821P.
 PR 14-JUL-2000; 2000US-218946P.
 XX
 FA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Triboley CM;
 PI Delemane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
 PI Wajia NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;
 PI Azimzai Y, Li Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
 PI Kallick DA;
 XX
 DR MPI; 2002-090437/12.
 DR P-PSDB; AAU74749.
 XX
 FT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful

in the diagnosis, treatment and prevention of gastrointestinal (e.g. gastritis), cardiovascular (e.g. atherosclerosis) and cell proliferative (e.g. cancer) disorders -

Claim 5; Page 166-167; 177bp; English.

The present invention relates to twenty one new human proteases, referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and polypeptides of the invention are useful in the diagnosis, treatment and prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and myocardial infarction, autoimmune/inflammatory e.g. acquired immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell proliferative e.g. cancer, developmental e.g. Duchenne and Becker muscular dystrophy, epithelial e.g. dermatitis, neurological e.g. epilepsy and Alzheimer's disease and reproductive e.g. infertility and endometriosis disorders. Numerous other examples of each disorder are given in the specification. The present nucleic acid sequence encodes the human protease PRTS-9 protein of the invention.

Sequence 3106 BP, 928 A, 633 C, 704 G, 841 T, 0 other;

Alignment Scores:
Pred. No.: 1.52e-204 Length: 3106
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 24 Gaps: 1

US-10-070-464-7 (1-360) x ABK12892 (1-3106)

QY 1 GUGUUAAPALAAAGSERAGLVALATHRPHVALLUENGUGLUPHASPARG 20
DB 932 GAAGAGATGCGAGATCGAGTCCCTTGTCTCCAGAAATTGATGGA 991
QY 21 TYRSEGLYTYRTPRCPYSPROLYSALAGUHRHPRROSEGLGLYLSILEU 40
DB 992 TATTCGCTANTGCTGCTGCCAAGCTGMAACATCCAGGCGGTAATTCCT 1051
QY 41 ARGILEUTYRGUGLUAASPNASPSGLUSEGLVALGULLEIHEISVALTHSER 60
DB 1052 AGAATTCTATATGAAGAAATGATGATCTGAGGGAATTTATTCATGTTACATCCCT 1111
QY 61 METLEUGUHRARXARXALASPSPHERGYTPROLYSTHGLYTHRALASPNRO 80
DB 1112 ATGTTGAAACAAGAGGCGCATTCCTGTTACCTTAAACAGGTAACGAATCCT 1171
QY 81 LYSVALTHRPHLYSMETSERGLULEMETILEASPALAGUGLYARGILEU 98
DB 1172 AAAGTCATTTTAAAGATGTCAGAAATATGATGATGCTGAAGAAAGATCATGATGTC 1231
QY 98 ----- 98
DB 1232 ATAGATAAGAACTAATTCACCTTTGAGATTCATTTGAAAGATTGAATATATGCC 1291
QY 98 ----- 98
DB 1292 AGAGCTGATGATCTCTGAGGAAATATGCTTGCTCATCTACTAGATCGCTCCAG 1351
QY 98 ----- 98
DB 1352 ACTCGCTACAGATAGTGTGATCTCACCTGAATATTATTCGCCAGTAGAAGATGATGT 1411
QY 98 ----- 98
DB 1412 ATGGAAGGACAGAGCTCATTTAGTCACTGCTGATTTCTGTGACGCCACTAATTAATCTAT 1471
QY 98 ----- 98
DB 1472 GAAGAAACAGACATCTGATTAATATTCATGATCATCTTTTCCCAAGT 1531
QY 98 ----- 98

DB 1532 CACGAAGAGAAATTGAGTTATTTTGGCTCTGATGCAAAAGAGTTCCGTCATTTA 1591
QY 98 ----- 98
DB 1592 TACAAATATCATCTATTTTAAAGAAAGCAATATATAAGATCCAGTGTGGCTGCT 1651
QY 98 ----- 98
DB 1652 GCTCCAGATGATTTCAAGTGTCTTCAAGAGAGATAGCAATTAACGATGATGATG 1711
QY 99 ----- 1711
DB 1712 GAAGTCTTGCCCGCATGATCTAATATCCAAAGTTATGAAAGTCAGAGGCTGTATAT 1771
QY 108 PHEGLUGLYTHRYSASPSPROLEUGLUIHSISLEUTYRVALSERTYRVALASN 127
DB 1772 TTGAAAGGACCAAGAACTCCCTTTAGAGCATCAGTGTAGTACGTAAAT 1831
QY 128 PROGLYGLUVALTHRARGLEUTHRASPARGLYTYRSEHISSERCYSYALLESGLIN 147
DB 1832 CCGAGAGGTCAGACAGCTGACCTGACCTGCTACTCACTTCTTGCTGATCAGTAG 1891
QY 148 HISCYASAPSPHEPHEILSERLYSTYRSEASNGLINYSAPSPROHISCYVALSERLEU 167
DB 1892 CACTGTACTCTTTATATAGATATAGTATACAGAAAGATCCACACTGTGTCCCTT 1951
QY 168 TYRILYSLEUSERSERPROGLUASPARPROTHRCYSLYSTHRLYSGLUPHETRALTHR 187
DB 1952 TACAAGCTATCAAGTCTGAAAGATGACCACTGCAAAACAAAGAAATTTTGGCCCAAC 2011
QY 188 ILELEUASPSERIALAGLYPROLEUPROASPHYRTHRPROPROGLUULEPHESEPHGLU 207
DB 2012 ATTTTGATGAGAGGCTCTTCTGACTATACCTCCAGAAATTTCTCTTTTGA 2071
QY 208 SERTHRGLYPHEHRLLEUTYRGLYMETLEUTYRILYSPROHISASPLEUGLINPROGLY 227
DB 2072 AGTACTAGATTTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2131
QY 228 LYSLYSTYRPROTHRVALLEUPHEILETYRGLYGLYPROGLINVALGILEUVALASN 247
DB 2132 AAGAAATATCCCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2191
QY 248 ARGPHLYSGLYVALLYSTYRPHARXLEUASPNHRLLEUASERLEUGLYTYRVALYAL 267
DB 2192 CGGTTTAAAGAGTCAGATATTTCCGCTTGAAATCCCTAGCCTCTTAGGTTAGTGCTT 2251
QY 268 VALVALLEASPAANARGLYSERCYSHISARGLYLEULYSPHEGLUGLYVALAPHELYS 287
DB 2252 GTAGGATATGACCAAGGAGATCCGTACCGAGAGGCTTAAATTTGAAGGCGCTTTAA 2311
QY 288 TYRILYSMETGLYGLINLEGLUULEASPARGLINVALGULYLEUGLINTYRLEUVALSER 307
DB 2312 TATAAAATGGGTCAAAATGAATGACATGATGATGATGATGATGATGATGATGATGATGAT 2371
QY 308 ARGTYRASPHEILEASPLEUASPARGLYVALGILYHEISGLYTPSPERTYRGLYLYTYR 327
DB 2372 CGATATGATTTTCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2431
QY 328 LEUSERLEUWETALAEUMECGLNARXSERASPILEPHEARGVALALALAGLYALA 347
DB 2432 CTCTCCCTGATGATGATTAATGACAGAGTCAGATATCTTCAAGGTTGCTATTTGCTGGGCGC 2491
QY 348 PROVALTHRLEUTPILIEPHETYRASPSTHRCGLYTYRTHR 360
DB 2492 CCACTACTCTGTGATCTCTATGATATCAAGATATCACG 2530
RESULT 7
AAC85694
ID AAC85694 standard; cDNA; 3120 BP.
XX AAC85694;
AC AAC85694;
XX

DT 29-JUN-2001 (first entry)

XX Nucleotide sequence of human DPP8.

DE Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;

KM dipeptidyl peptidase; DPP8; T cell; cleavage; diarrhoea;

KM growth hormone deficiency; glucose level; mucosal regeneration;

KM non-insulin dependent diabetes mellitus; glucose intolerance;

KM immunosuppression; ss.

XX Homo sapiens.

OS

XX Key Location/Qualifiers

FT CDS 214..2862

FT /tag= a

FT /product= "Human DPP8"

XX W0200119666-A1.

XX 22-MAR-2001.

XX 11-SEP-2000; 2000WO-AU01085.

XX 10-SEP-1999; 99AU-0002762.

PR 18-FEB-2000; 2000AU-0005709.

XX (UNSY) UNIV SYDNEY.

PA Abbott CA, Gorell MD;

PI WPI: 2001-261520/29.

XX P-PSDB; AAB47187.

DR

XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving

PT substrates, identifying inhibitors of DPP8 catalytic activity which

PT have therapeutic uses, and for detecting activated T cells

XX

PS Claim 16; Fig 2; 78pp; English.

XX

CC This sequence encodes human dipeptidyl aminopeptidase (DPP8). DPP8

CC has substrate specificity for H-Ala-Pro-pNA, H-Gly-Pro-pNA and

CC H-Arg-Pro-pNA. Therefore, it is a prolyl oligopeptidase and a

CC dipeptidyl peptidase, because it is capable of hydrolysing the

CC peptide bond C-terminal to Pro in each of these compounds. DPP8

CC is homologous with human DPP8. DPP8 is useful for cleaving a

CC substrate, and for detecting an activated T cell which involves

CC measuring the level of DPP8 gene expression in a T cell. The level

CC of DPP8 expression is detected by detecting the amount of DPP8 RNA

CC in the cell. It is also useful for identifying a molecule capable

CC of inhibiting the cleavage of the substrate by DPP8. Molecules

CC identified as inhibiting DPP8 catalytic activity may be useful for

CC treating diarrhoea, growth hormone deficiency, lowering glucose levels

CC in non-insulin dependent diabetes mellitus and other disorders

CC involving glucose intolerance, enhancing mucosal regeneration and

CC as immunosuppressants.

XX

SQ Sequence 3120 BP, 936 A; 637 C; 706 G; 841 T; 0 other;

XX

Alignment Scores:

Pred. No.: 1,536-204 Length: 3120

Score: 1836.50 Matches: 360

Percent Similarity: 67.54% Conservative: 0

Best Local Similarity: 67.54% Mismatches: 0

Query Match: 95.01% Indels: 173

DB: 22 Gaps: 1

US-10-070-464-7 (1-360) x AAC85694 (1-3120)

QY 1 GIUGLUSPALAAGSERIALAGLYVALAAlaThrPheValLeuGInGluGluPheAspArg 20

DB 943 GAAGAATAATGCGCAATCAGCTGAGTGGCTACCTTTGTTCTCCCAAGAAATTTGATAGA 1002

QY 21 TysSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGlyLeuLeu 40

DB 1003 TATTCGCTAATGTGTGTGTCGAAAGCTGAAACCACTCCAGTGTGTGTAATTTCTT 1062

QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60

DB 1063 AGATTCTATATGAAAGAAATGATGAAATCTGAGTGGAAATTTATTCATGTTACATCCCT 1122

QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80

DB 1123 ATGTGGAAACAAAGAGGCGAGATTCAATCCGTATCCTAAACAGGTACGCAAACTCT 1182

QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluValArgIleIle----- 98

DB 1183 AAAGTCACTTTTAAAGATGTGAGAAATATGATGATGCTGGAAGAGGATCATGATGTC 1242

QY 98 ----- 98

DB 1243 ATAGATAAGAACTAATTCAACTTTTGAGATTCTATTGGAAGAGTTGAATATATTCGC 1302

QY 98 ----- 98

DB 1303 AGAGCTGATGAGACTCCTGAGGAAATATGCTTGTCATCTAGATGCTCCGAG 1362

QY 98 ----- 98

DB 1363 ACTGCCTACAGATAGTGTGATCTCACTGAAATTATTTATCCAGTAGAAGATGATGTT 1422

QY 98 ----- 98

DB 1423 ATGGAAGAGCAGAGACTCATGAGTCACTGCTGCTGATTTCTGAGCCCATATATCTAT 1482

QY 98 ----- 98

DB 1483 GAAGAAACAAACAGACATCTGGATTAATATCCATGACATCTTTCATGTTTCCCAAGT 1542

QY 98 ----- 98

DB 1543 CACGAAGAGAAATGAGTTATTTTGTGCTCTGAATGCAAAAGAGTTTCGTCATTTA 1602

QY 98 ----- 98

DB 1603 TACAAATTTACATCTATTTTAAAGAAAGCAAAATATAACATCCAGTGGGTGCT 1662

QY 98 ----- 98

DB 1663 GCTCAAGATGATTTCAAGTGTCTATCAAGAGAGATAGCAATTACCAAGTGGTGAATGG 1722

QY 99 -----ValaspgluValArgArgLeuValTyr 107

DB 1723 GAAGTTCTTGCCCGCGCATGATCTAATATCCAAAGTTGATGAGTCAGAGGCTGTATAT 1782

QY 108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127

DB 1783 TTTTAAAGCACCAAGACTCCCTTTAGAGATCACTGTACGTACGTAGTACGTAAAT 1842

QY 128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGln 147

DB 1843 CCTGAGAGAGTGAACAAAGCTGACGACCGTGTACTACATCTCTGTCGATCAGTCAG 1902

QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167

DB 1903 CACTGTACTCTTTTAATAGATAGTATAGTATACCAAGAGATCCACTGTGTCTCTT 1962

QY 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 187

DB 1963 TACAGCTATCACTGCTGAAAGATGACCACTGCAAAACAAAGAAATTTGGGCCACC 2022

QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207

DB 2023 ATTTTGATTCAGAGGTCTCTTCTGACTATCTCTCCAGAAATTTTCTTTTGAA 2082

QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227

Db 2083 AGTACTACGATTACATTTGATGGATGCTCTCAAGGCTCATGATCTACAGCCTGGA 2142
 QY 228 LysLysIYrProThVAlleuPheIleTyrgIyglYProGlnValGlnleuValAsnAsn 247
 Db 2143 AAGAAATATCTTACTGCTGTGCTTATATATGATGCTCTCAAGGCTCATGATCTACAGCCTGGA 2202
 QY 248 ArgPheLysGlyValLysIYrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 Db 2203 CGGTTTAAAGAGTCAAGTATTTCCGCTTGATACCTTACCTCTCTAGGTTATGTT 2262
 QY 268 ValValIleAspAsnArgIYSerCysHisArgGlyLeuLysPheGlnGlyAlaPheLys 287
 Db 2263 GTAGGATAGACAAACAGGGGATCCGTGCACGAGGGCTTAATTTGAAGGCCCTTTAA 2322
 QY 288 TyrLysMetGlyGlnIleGlnIleAspAsnGlnValGlnGlyLeuGlnIYrLeuAlaSer 307
 Db 2323 TATAAAATGGGTCAATATGAAATATGACATCAGGTGAGGAGGACTCCAAATATCTACCTCT 2382
 QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyIYr 327
 Db 2383 CGATATGATTCATTTGACTTACCTGCTGTGGGATCCAGGCTGTGCTCTATGAGAGATAC 2442
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 Db 2443 CTCTCCCTGATGGCATTAATGACAGAGGTCAAGATATCTTCAAGGCTGTGCTATGCTGGGCC 2502
 QY 348 ProValThrLeuThrIlePheTyrAspThrGlyTyrThr 360
 Db 2503 CCAGTCACCTGTGTGATCTTCTATGATACAGGATACAGC 2541
 RESULT 8
 AAD38956
 ID AAD38956 standard; cDNA, 3120 BP.
 AC AAD38956;
 DT 23-SEP-2002 (first entry)
 DE Human dipeptidyl peptidase 8 (DPP8) cDNA.
 KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
 KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
 KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
 KW antiviral; enzyme; gene; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 214..2862
 FT /tag= a
 FT /product= "Human DPP8 protein"
 FN WO200234900-A1.
 PD 02-MAY-2002.
 PF 29-OCT-2001; 2001WO-AU01388.
 XX 27-OCT-2000; 2000AU-0001078.
 PA (UNSY) UNIV SYDNEY.
 PI Abbott CA, Gorrell MD;
 DR WPI: 2002-454646/48.
 DR P-PSDB; AAE24170.
 PT New dipeptidyl peptidase (DPP) peptides, useful for screening
 PT inhibitors of DPP catalytic activity, which may be employed to treat
 PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 PT rejection and HIV infection -
 PS Example; Fig 1; 91pp; English.

XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
 CC polymucleotides encoding such proteins. The DPP peptides are useful for
 CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
 CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 CC rejection and HIV (human immuno deficiency virus) infection. The present
 CC sequence is human DPP8 cDNA.
 SQ Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;
 Alignment Scores:
 Pred. No.: 1,53e-204 Length: 3120
 Score: 1836.50 Matches: 360
 Percent Similarity: 67.54% Conservative: 0
 Best Local Similarity: 67.54% Mismatches: 173
 Query Match: 95.01% Indels: 1
 DB: 24 Gaps: 1
 US-10-070-464-7 (1-360) x AAD38956 (1-3120)
 QY 1 GlnGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArg 20
 Db 943 GAAGAAAGTGCAGATCAGCTGAGATCGCTACCTTGTCTCCAGAAAGAAATTTGATGA 1002
 QY 21 TyrSerGlyTyrTrpCysProLysAlaGlnThrThrProSerGlyGlyLysIleLeu 40
 Db 1003 TATTCGGCTATTGGTGTGTGTCCTCAAGACTAAACACTCCAGTGGTGTAATTCCT 1062
 QY 41 ArgIleLeuTyrGlnGluAsnAspGluSerGluValGlnIleHisValThrSerPro 60
 Db 1063 AGAATTCATATGAAGAAATGATGAATCTAGAGTGAATTAATTCATGTTACATCCCT 1122
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrAlaAsnPro 80
 Db 1123 ATGTTGAAACAGAGGCGAGATTCATTCCTTAATCTCAAAACAGTACAGCAATCCT 1182
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGlnGlyArgIleIle 98
 Db 1183 AAAGTCACCTTTAAGATGTCAGAAATATGATGATGCTGTAAGAGAGATCATTAATGTC 1242
 QY 98 ----- 98
 Db 1243 ATGATTAAGAACTAATCAACCTTTGAGATTCATTTGAAGAGATTGAATATATGCGC 1302
 QY 98 ----- 98
 Db 1303 AGAGCTGATGACTCTCTGAGGAAATATGCTTGTCATCTCATAGATGCTCCAG 1362
 QY 98 ----- 98
 Db 1363 ACTGGCTACGATGATGTTGATCTCACCTGAATTTATTCCTCCAGTGAAGATGATGTT 1422
 QY 98 ----- 98
 Db 1423 ATGGAAGGACAGACTCATTTAGTCAAGTGCCTGATCTGTGAGCGCACATTAATCTAT 1482
 QY 98 ----- 98
 Db 1483 GAAGAAACACAGACATCTGATTAATATTCATGATCTTTATGTTTCCCAAGT 1542
 QY 98 ----- 98
 Db 1543 CACGAAAGGAATTTAGTTTATTTTGGCTCTGAATGCAAAACAGTTTCCGTCATTTA 1602
 QY 98 ----- 98
 Db 1603 TACAATAATTCATCTATTTTAAGAAACAAATATAACGATCCAGTGGGCTGCT 1662
 QY 98 ----- 98
 Db 1663 GCTCCAAAGTATTTCAAGTGTCTTATCAAAAGAGAGATGCAATTTACAGTGGTGAATGG 1722
 QY 99 -----ValAspGluValArgArgLeuValTyr 107

```

Db      1723 GAAGTCTTGGCCGGCATGATCTAAATATCCAAAGTTGAAGAACTGGAAGGCTGGATATAT
Qy      108 pHeGluGlyThrLysAspSerProLeuGluHisLysLeuValValSerTyValAsn 127
Db      1783 TTTGAAGGACCAAAAGACTCCCTTTAGAGCATCACTGACTAGTACGTTACGTAAT 1842
Qy      128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
Db      1843 CCGGAGAGGTGACAAAGGCTGACTGACCGGGCTACTCATCTTGCTGCACACTCAG 1902
Qy      148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db      1903 CACTGTGACTCTTTTAAAGTAAGTATAGTACCAAGAAATCCACATGCTGTGCTCT 1962
Qy      168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 187
Db      1963 TCAAGCTATCAAGTCTCGAAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCACC 2022
Qy      188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db      2023 ATTTGGATTGACAGAGTCTCTCTCTGACTATATCTCCAGAAATTTTCTTTTGAA 2082
Qy      208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db      2083 AGTACTACTGGATTTCATTTGATGAGAGTCTTACAGGCTCATATCTACAGCTTGA 2142
Qy      228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 247
Db      2143 AAGAAATATCTPACTGTGCTGTTCAATATAGTGTCCTCAGGTGACAGTTGGTAATAT 2202
Qy      248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db      2203 CGGTTTAAAGAGTCAAGATATTCCTGCTGAAATACCTGACCTTAGGTATATGCGTT 2262
Qy      268 ValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluAlaPheLys 287
Db      2263 GATGATATGACAAACAGGGGATCTCTTACCGAGGCTTAATTTAAAGCCCTTTAA 2322
Qy      288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db      2323 TTTAAATGGTCAATATGAATTTGACGATCAGGTGAAAGAACTCCAAATATCTAGCTTCT 2382
Qy      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyGlyTyr 327
Db      2383 CCAATATGATTTCACTTACTTAAATCGTGGGATCCAGCGTGGTCTTATGAGGATAC 2442
Qy      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db      2443 CTCTCCCTATGCGATTAAATGACAGAGTACAGATATCTTCAAGGTTCTATTGCTGAGGCC 2502
Qy      348 ProValThrLeuThrIlePheTyrAspThrGlyTyrThr 360
Db      2503 CCAGTCACTCTGTGATCTTCTATGATACAGGATACAGC 2541

```

```

RESULT 9
AAH99934
ID AAH99934 standard; cDNA; 3143 BP.
AC AAH99934;
XX
XX
DT 12-APR-2002 (first entry)
XX
XX cDNA encoding 21953 human prolyl oligopeptidase.
XX
XX 21953 prolyl oligopeptidase; human; proline; endopeptidase;
XX cancer; cardiovascular disease; autoimmune disease; atopic allergy;
XX neuronal disorder; vascular disorder; prostate disorder; cytostatic;
XX antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
XX diabetes mellitus; arthritis; multiple sclerosis; asthma;
XX Grave's disease; neuronal disorder; demyelinating disease; ss.
OS Homo sapiens.

```

```

XX Key Location/Qualifiers
FH CDS 229..2877
FT /tag= a
FT /product= "21953 prolyl oligopeptidase"
FT /note= "This region is specifically claimed in
FT 0 ' claim 2"
PN WO200179473-A2.
XX
XX 25-OCT-2001.
PD
XX
XX 11-APR-2001, 2001WO-US40483.
XX
XX 18-APR-2000, 2000US-197508P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PI Meyers RA, Williamson M;
XX
XX MPI; 2002-034353/04.
XX P-PSDB; AAG78415.
XX
XX New polypeptides 21953, member of human prolyl oligopeptidase family,
XX useful as diagnostic targets and therapeutic agents for controlling
XX cancer, lymphoma and leukemia
XX
XX Claim 7, Page 100-102; 121p; English.
XX
XX This invention relates to an isolated 21953 human prolyl
XX oligopeptidase. Which is cytostatic, antidiabetic, antiarthritic,
XX neuroprotective, antihypoid, dermatological, antipsoaritic,
XX antiasthmatic, ophthalmological, antiinflammatory, nootropic,
XX antiparkinsonian, anticonvulsant, gynaecological, vasorelaxant,
XX antianginal, cardiant, antiatherosclerotic, anorectic and
XX metabolic in its action. Uses include gene therapy, expression or
XX compound of 21953 protein modulator, it is useful for identifying a
XX activity which binds to it and can be used in preventing, treating
XX or detecting a cellular proliferative or differentiative disorder.
XX The 21953 molecules can act as novel diagnostic targets and therapeutic
XX agents for controlling disorders associated with the aberrant activity
XX or degradation of peptide hormones e.g., disorders associated with cell
XX differentiation and proliferation such as cancer, immune function,
XX reproductive, neurological and cardiovascular function. The 21953
XX molecules are thus useful for treating and preventing cellular
XX proliferative and differentiative disorders, haematopoietic neoplastic
XX disorders, immune disorders such as autoimmune diseases, diabetes
XX mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
XX neuronal disorders, demyelinating diseases, vascular disorders and
XX metabolism or pain disorders. This sequence represents the cDNA
XX encoding sequence of 21953 human prolyl oligopeptidase.
XX
XX Sequence 3143 BP; 943 A; 644 C; 712 G; 844 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.54e-204 Length: 3143
XX Score: 1636.50 Matches: 360
XX Percent Similarity: 67.54% Conservatve: 0
XX Best Local Similarity: 67.54% Mismatches: 0
XX Query Match: 95.01% Indels: 173
XX DB: 24 Gaps: 1
XX
XX US-10-070-464-7 (1-360) X AAH99934 (1-3143)
Qy 1 GluLysAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
Db 958 GAAGAAAGATCCAGATACACTGAGTGGTCTACTTTGTTCTCCAAAGAAATTTGATAGA 1017
Qy 21 TyrSerGlyTyrTyrTyrPyrCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
Db 1018 TATTTGCTATTTGGTGGTGTCCAAAGCTGAACCAACTCCAGTGGTGAATTTCTT 1077
Qy 41 ArgIleLeuTyrGluGluLysAspGluSerGluValGluIleIleHisValThrSerPro 60

```

```

Db      1078 AGAATTCATATAGAAATATGATGATCGAGGTGGAATTAATCATGTTACATCCCT 1137
Qy      61 MetLeuGluThrArgArgIAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db      1138 ATGTGGAAACAAAGAGGCGAGATTCATCCGTTATCTAAACAGGTACAGCAAAATCCT 1197
Qy      81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluIleIle----- 98
Db      1198 AAGATCATTTTAAGATGTCAGAAATATGATGATCTGAGGAGAGATCATGATGTC 1257
Qy      98 ----- 98
Db      1258 ATGATATAGGAACTAATTCACACCTTTGAGATTCATTTGAGAGAGTTGATATATGCC 1317
Qy      98 ----- 98
Db      1318 AGAGCTGATGAGCTCCTGAGGAAATATAGCTTGTCATCTACTAGATCGCTCCAG 1377
Qy      98 ----- 98
Db      1378 ACTGCGCTGAGATGATTGTTGATCTCACTGAATTATTATCCAGTGAAGATGATGTT 1437
Qy      98 ----- 98
Db      1438 ATGGAAGGAGAGACTCATTTAGTCACTGCGCTGATTCGTGAGAGCCACTAATATCTAT 1497
Qy      98 ----- 98
Db      1498 GAAGAAACAAAGACATCTGATTAATATCCATGATCATCTTCAATGTTTTCCCAAGT 1557
Qy      98 ----- 98
Db      1558 CACGAAGAGAAATTTGATTATTTTTCCTCTGATGCAAAACAGSTTCCGTCATTTA 1617
Qy      98 ----- 98
Db      1618 TACAAATATCATCTATTTTAAAGAAAGCAATATATAAGATCCAGTGGGCTGCTCT 1677
Qy      98 ----- 98
Db      1678 GCTCCAGATGATTTCAAGTGTCTATCAAAAGAGATACCAATTCAGTGTGAATGG 1737
Qy      99 ----- ValAspGluValArgArgLeuValTyr 107
Db      1738 GAAGTCTTGCCCGCATGATCTAATATCCAGTTATGAGTCAGAGAGCGCTGATAT 1797
Qy      108 PheGluGlyThrLysAspSerProLeuGluHisIleuTyrValValSerTyrValAsn 127
Db      1798 TTGGAAGGACCAAGAGCTCCCTTTAGAGATCACCCTGTAACGTAAGTCAAGTAAAT 1857
Qy      128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 147
Db      1858 CCTGGAAGGAGTACAAAGCTGACTGACCGGTGCTACTCACTCACTTCTGATCAGTCAG 1917
Qy      148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db      1918 CACTGATCTTCTTTATAGTAAGTATAGTAACAGAAAGATCCACACTGATGTCCTT 1977
Qy      168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheMetPalThr 187
Db      1978 TACAAGCTATCAACTCTGAAAGATGACCAACTTGCAAAACAAAGGAATTTGGGCCACC 2037
Qy      188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db      2038 ATTTTGATTCAGAGAGTCTCTTCTGATCACTACTCTCCAGAAATTTTCTTTTGA 2097
Qy      208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db      2098 AGTACTACTGATTTATCATGTATGATGATGCTCAACAGCTCATGATCAACACTGGA 2157
Qy      228 LysLysTyrProThrValIleuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247

```

```

Db      2158 AAGAAATATCTACTGCTGCTCATATATGATGTCCTCAGAGTGCAGTTGTGAATAT 2217
Qy      248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db      2218 CGGTTTAAAGAGATCAAGATATTCGCTTGAATAACCTTAGCCTCTTAGGTATATGGTT 2277
Qy      268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db      2278 GTAGTATAGACAAACAGGGATCTCTCACCAGGGCTTAAATTTGAAGCGCTTTAA 2337
Qy      288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db      2338 TATTAATAGGCTCAATTAAGAAATTCAGATCAGGTGGAAGAGACTCCAAATATCTAGCTTCT 2397
Qy      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyrGlyGlyTyr 327
Db      2398 CGATATGATTTTCATTACATTAGATCGATGGGATCCACGGCTGCTTATGAGGATAC 2457
Qy      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db      2458 CTCTCCCTGATGCGATTTATGACAGAGTCAAGATATCTTCAGGGTGTCTATGCTGGGGCC 2517
Qy      348 ProValThrLeuThrPheTyrPheTyrAspThrGlyTyrThr 360
Db      2518 CCAGTCATCTCTGATGATCTTCTATGATACAGATACACG 2556

RESULT 10
ABK83327 standard; cDNA; 4829 BP.
XX
AC ABK83327;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPP-1 splice variant #3.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001MO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Jurnen J;
XX
DR WPI; 2002-444178/47.
XX
DR P-PSDB; AB661596.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PS Disclosure; Page 65-66; 113pp; English.
XX
XX The present invention relates to the isolation of novel human serine
XX proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
XX proteases (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
XX and nucleic acids encoding them are useful for treating infections
XX such as fungal, bacterial, protozoan and viral infections, particularly

```

CC infectious caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenal), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABRK3332-ABRK3343 encode human DPR proteins.

Sequence 4829 BP; 1466 A; 886 C; 1017 G; 1460 T; 0 other; SQ

Alignment Scores:

Ered. No.:	6.56e-202	Length:	482
Score:	1816.50	Matches:	360
Percent Similarity:	67.29%	Conservative:	0
Best local Similarity:	67.29%	Mismatches:	0
Query Match:	93.97%	Indels:	175
DB:	24	Gaps:	1

US-10-070-464-7 (1-360) x ABK83327 (1-4829)

QY	1	lhuGluaSpAlaargSerAlaGlyValAlaThrPheValIleuGlnGluPheaspArg	20
Db	943	GAAGAGATGCCAGATCAGCTGGAGTCCCTACCTTGTCTCCAAAGAAATTGATAGA	1002
QY	21	TySerGlyTyTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu	40
Db	1003	TATTCGTGCTATTTGGTGTGTCGTCCAAACCTGAACAACTCCCAAGTGGTGAATTCCTT	1063
QY	41	ArgIleLeuTyArgGluGluAsnaspGluSerGluValGluIleIleHisValThrSerPro	60
Db	1063	AGAAATTCTATATGAAGAAATATGATGAATCTGAGGGGAAATTAATTCATGTTACATCCCT	1122
QY	61	MetIleuGluThrArgArgAlaAspSerPheArgTyTrpProIleThrGlyThrAlaAsnPro	80
Db	1123	ATGTTGGAAACAAGGAGGCGAGATTCATTCGTTATCTTAAACACGATACGCAATTCCT	1182
QY	81	LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle-----	98
Db	1183	AAAGTCACTTTTAAGATGTCAGAAATATATGATTGATGCTGAAGAGAGATCATAGATGTC	1242
QY	98	-----	98
Db	1243	ATAGATGAAGAACTAATTCAACCTTTGAGATTCATTGTAAGAGGTTGAATATATTGCC	1302
QY	98	-----	98
Db	1303	AGAGCTGATGACTCTCTGAGGAAATATGCTTGGTCACTCACTAGATGCTGCCAG	1362
QY	98	-----	98
Db	1363	ACTGCGCTACAGATATGTGTTGATCTCACCTGAATTAATTATTCACAGTAAAGATGAGTT	1422
QY	98	-----	98
Db	1423	ATGGAAGGCAAGAGACTCATGAGTCAATGCGCTGATTTCTGAGCGCCACTAATTATCTAT	1482
QY	98	-----	98
Db	1483	GAAGAAACACAGACATCTGATTAATATCATGACATCTTTCATGTGTTTTCCCAAGT	1542
QY	98	-----	98
Db	1543	CACGAGAGAGAAATGAGTTATTTTGGCTCTGTGATGCAAAACAGGTTCCGTCATTTA	1602
QY	98	-----	98
Db	1603	TACAAATATCATCTATTTTAAAGGAAAGCAATATTAAGATCCAGTGTGGGCTGCT	1662
QY	98	-----	98
Db	1663	GCTCCAGATGATTTCAAGTGTCTATCAAAAGAGAGATGACAAATTACCAAGTGTGAATGG	1722

QY	99	-----ValAspGluValAlaGlyLeuValTyr	107
Db	1723	GAAGTCTTGACCGGACATGATCTAATAATCCAAAGTTATATGAAGTCAAGAGGCTGGTATAT	1782
QY	108	PheGluGlyThrLysAspSerProLeuGluHisIleuTyrValIleSerTyrValAsn	127
Db	1783	TTTGAGGACACCAAGACTCCCTCTTTAGAGCATCACCTGTAAGTCAGTTACGTAAAT	1842
QY	128	ProGlyGluValThrArgLeuThrAspArgLysTyrSerHisSerCysCysIleSerGln	147
Db	1843	CCTGAGAGGAGTGACAAAGGCTGACTGACCGGTGGCTACTCACATTTCTGTGTCATCAGTCAG	1902
QY	148	HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValIleLeu	167
Db	1903	CACCTGTGCTCTTTATATGATGATATGATATGATACCAAGAAATCCACACTGTGTCTCCCTT	1962
QY	168	TyrLysLeuSerSerProGluAspAspProHisCysLysThrLysGluPheThrAlaThr	187
Db	1963	TACAAAGCTATCAAGTCTGAAGATGACCACACTTGCAAAACAAAGAAATTTGGGCCACC	2022
QY	188	IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu	207
Db	2023	ATTTTGATTCAGACAGGCTCTCTCTCTGACTATATCTCTCCAGAAATTTCTCTTTGAA	2082
QY	208	SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly	227
Db	2083	AGTACTACTGATTTTACATTTGATTAAGGATGGCTCAACAGCTCTCATGATCTTACACTTGA	2142
QY	228	LysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnLeuValAsn	247
Db	2143	AAGAAATATCTCTACTGCTGCTTTCAATATATGATGGTCTCTCAGGGTCAGTTGGTGAATA	2202
QY	247	snArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal	267
Db	2283	ATCGTTTAAAGAGTCAAGTATTTCCGCTTGAAATCCCTAGCCTCTTACGTTATATGG	2262
QY	267	AlValValIleAspAsnArgLysCysHisArgGlyLeuLysPheGluGlyAlaPheL	287
Db	2263	TTGTATGATATGACAAACAGGGAGTCCGTCACCCAGGGCTTAAATTTGAAAGCCGCTTTA	2322
QY	287	LysTyrLysMetGlnGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnIleTyrLeuAla	307
Db	2323	AATATATAATGGGTCAATATAGAAATGACGATCAGGTGAAGAGACTCCAAATATCTAGCTT	2382
QY	307	erArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyGlyT	327
Db	2383	CTCGATTATGATTTATTGACTTAACTGATGCTGTGGCATTCACGCTGTGCTCATGGAGAT	2442
QY	327	YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly	347
Db	2443	ACCTCTCCCTATGGCATTTAATGACAGGTCAGATATCTTCAGGGGTGTCTATTGCTGGGG	2502
QY	347	IaProValThrLeuThrIlePheTyrAspThrGlyTyrThr	360
Db	2503	CCCCAGTCACCTGTGGATCTTCTATATGATACAGATACAG	2543
RESULT 11			
AAH15009 standard; cDNA; 2161 BP.			
AC	AAH15009;		
XX	26-JUN-2001	(first entry)	
DT	Human cDNA sequence SEQ ID NO:12963.		
XX			
DE	Human, primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	EP1074617-A2.		

PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 XX
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX
 PS Claim 8; SEQ ID 12963; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2161 BP; 662 A; 429 C; 452 G; 618 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 1.38e-168 Length: 2161
 Score: 1528.00 Matches: 308
 Percent Similarity: 57.79% Conservative: 0
 Best Local Similarity: 57.79% Mismatches: 1
 Query Match: 79.05% Indels: 224
 DB: 22 Gaps: 2
 US-10-070-464-7 (1-360) x AAH15009 (1-2161)
 QY 1 GUGUASPAlAARgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
 DB 148 GAGAGAGATGCCAGACACCTGGAGTCGCTACCTTGTCTCCAAAGAAATTTGATGAA 207
 QY 21 TYSerGlyTyrTriPTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
 DB 208 TATTCGGCTATTGGTGGTGTCTCCAAAGCTGAACCACTCCAGTGGTGAATAATTCCT 267
 QY 41 ArgIleLeuTyrGlnGluAsnAspGluSerGluValGlnIleIleHisValThrSerPro 60
 DB 268 ACAATTCATATGAAAGAAATGATGATCTGAGGTGGAATATTATTCATGTCATCCCT 327
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80

DB 328 ATGTTGAAAACAAGAGGGCAGATTCAATCCGTTATCTTAAAAACAGGTACAGCAATCCT 387
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
 DB 388 AAAGTCACCTTTTAAAGTGTGCAAGAAATAAATGATTGATGTGGAAGTAAAGATTCATAGATGTC 447
 QY 98 ----- 98
 DB 448 ATAGATMAAGAACTAATTCACCTTTTGATTCATTATTGAAAGAGTGAATATATATGCCC 507
 QY 98 ----- 98
 DB 508 AAGAGCTGATGAGACTCTCGAGGGAAATATGCTTGCTGCATCCTACTAGATGCTCCAG 567
 QY 98 ----- 98
 DB 568 ACTCGCTACAGATAGTGTGTGATCTCACCTGATATTATTTATCCAGTAGAAGATGATGTT 627
 QY 98 ----- 98
 DB 628 ATGAAAAGCAGAGACTCATTTGATGATGCTGCTGATTCGTGAGCCACTAATTATCTAT 687
 QY 98 ----- 98
 DB 688 GAAAGAAACAACAGACATCTGATMAAATATCATGACATCTTCATGTTTTCCTCCAAAGT 747
 QY 98 ----- 98
 DB 748 CACGAGAGAGAAATGAGTTTATTTTGGCTGTGATGCAAAAACAGTTTCCGTCAATTGA 807
 QY 98 ----- 98
 DB 808 TACAAATATTCATCTATTTTTAAAGAAAGCAATATTAACGATCCAGTGGTGGCTGCT 867
 QY 98 ----- 98
 DB 868 GCTCCAGAGATTTCAAGTGTCTTCAAAAGAGAGATGCAATTACAGTGGTGAATGG 927
 QY 99 -----ValAspGluValArgProLeuValTyr 107
 DB 928 GAAGTTCTGGCCGGCAGATGATCTATATCAAGTTGAGGAGAGTGAAGGCTGGTATAT 987
 QY 108 PheGluGlyThrLysAspSerProLeuGluHisIleLysLeuTyrValValSerTyrValAsn 127
 DB 988 TTGGAAGGACCAAAAGATCTCCCTTTAGAGCATCTACTGTACTGATGTCAGTTACGAAAT 1047
 QY 128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
 DB 1048 CCTGGAGAGGTGACAAGGCTGACTACCGTACTACTACATTTCTTGCTGCATCAGTCAAG 1107
 QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
 DB 1108 CACTGTGACTTCTTTATAGTATGATGATTAACCAAGAAATCCACACTGTGTGCTCTT 1167
 QY 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluIlePheTrpAlaThr 187
 DB 1168 TCAACAGCTATCAAGTCCGTAAGATGACCAACTTGCAAAACAAAGAAATTTTGGGCAACC 1227
 QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
 DB 1228 ATTTTGATTCAGCAGAGGCTCTCTCTGCTACTACTCCCGAAGAAATTTCTTTTGAA 1287
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 DB 1288 AGTACTACTGATGATTAATGATGATGATGATGCTTCAAGCCCTCAAGATATTAAGCGCTGGA 1347
 QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
 DB 1348 AAGAAATATCCACAGTGTGTTATATATGTTGCTCCAGGTGCAAGTGGTGAATAT 1407
 QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 DB 1408 CGGTTTAAAGAGTCAAGATATTCGCTTGAAATACCTTAGCTCTAGTTATATGCTT 1467


```

OY 268 ValVal11leAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaIlePheLys 287
DB 1468 GTAGTGATGACCAACAGGGGATCCCTGTCACCCAGGGCTTAAATTGGAAGGCCCTTTAA 1527
OY 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnIleTyrIleuLaser 307
DB 1528 TATTAATAAG----- 1536
OY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 327
DB 1536 ----- 1536
OY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
DB 1537 ----- GTTGTATTTCCTGGGCC 1554
OY 348 ProValThrLeuThrPhePheTyrAspThrGlyTyrThr 360
DB 1555 CCACTGACTCTGTGATCTTCTATGATACAGGATACACG 1593
RESULT 12
AAD23843
ID AAD23843 standard; cDNA; 2510 BP.
XX
AC AAD23843;
XX
DT 07-MAR-2002 (first entry)
XX
DE Human protease PRTS-2 cDNA.
XX
KW Human; protease; PRTS-2; tranquilizer; gene therapy; vaccine; allergy;
KW infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;
KW atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;
KW gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;
KW epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;
KW hypertension; neurological disorder; Parkinson's disease; drug screening;
KW cardiast; cell proliferative disorder; multiple sclerosis; osteoporosis;
KW diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;
KW autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;
KW developmental disorder; reproductive disorder; infertility; diarrhoea;
KW dementia; acidosis; cataract; gynaeocomastia; epilepsy; jaundice; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 616..2358
FT /tag= a
FT /product= "Human protease PRTS-2 protein"
XX
PN WO200183775-A2.
XX
PD 08-NOV-2001.
XX
PF 04-MAY-2001; 2001WO-US14651.
XX
PR 04-MAY-2000; 2000US-202082P.
PR 11-MAY-2000; 2000US-203566P.
PR 17-MAY-2000; 2000US-205803P.
PR 25-MAY-2000; 2000US-207477P.
PR 01-JUN-2000; 2000US-209402P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Deleageane AM, Lal P, Hafalia A, Patterson C, Walia NK, Kearney L,
PI Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS,
PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM,
PI Reddy R, Yue H, Tang YT;
XX
DR WPI; 2002-034518/04.
DR P-PSDB; AAE14337.
XX
PT Novel human proteases and polynucleotides encoding the proteases,

```

PT useful for treating, diagnosing or preventing cell proliferative,
PT cardiovascular, autoimmune/inflammatory, neurological and developmental
PT disorders -

PS Claim 5; Page 139-140; 151pp; English.

CC The invention relates to human proteases (PRTS-14) and its corresponding
CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for
CC the diagnosis, treatment and prevention of disorders associated with
CC increased or decreased expression of PRTS. Examples of such disorders
CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,
CC hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders
CC (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,
CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,
CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and
CC viral, bacterial, fungal, parasitic, protozoal and helminthic
CC infections); cardiovascular disorders (myocardial infarction, ischemic
CC heart disease and hypertension); neurological disorders (epilepsy,
CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,
CC Parkinson's disease, stroke, mental disorders including mood, anxiety
CC and seasonal affective disorder and prion diseases); gastrointestinal
CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);
CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,
CC alopecia, scabies, insect bites and urticaria); reproductive disorder
CC (infertility, disruption of estrous and menstrual cycle and
CC gynaeocomastia); and developmental disorders (renal tubular acidosis,
CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).
CC PRTS DNA is also in useful in gene therapy. PRTS and its immunogenic
CC fragments are useful for screening libraries of compounds in several drug
CC screening assays. The present sequence is human protease PRTS-2 cDNA.

XX SQ Sequence 2510 BP; 777 A; 494 C; 527 G; 712 T; 0 other;

Alignment Scores:

Pred. No.:	8 79e-168	Length:	2510
Score:	1522.00	Matches:	309
Percent Similarity:	58.08%	Conservative:	0
Best Local Similarity:	58.08%	Mismatches:	0
Query Match:	78.74%	Indels:	223
DB:	24	Gaps:	2

US-10-070-464-7 (1-360) x AAD23843 (1-2510)

```

OY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArg 20
DB 609 GAAGAAATGCCAGATCGCTGAGTGCCTACCTTGTCTCCAGAAAGAAATTGATGGA 668
OY 21 TyrSerGlyTyrTPTPTPCysProLysAlaGluThrThrProSerGlyGlyLysIleu 40
DB 669 TATTCTGCTATTGGTGTGTGCCAAAGCTGAAACACCTCCACAGTGCGTAAATTCTT 728
OY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleHisValThrSerPro 60
DB 729 AGAATTCATATGAGAAAGATGATGAATCTAGGTGAAATTTATTCATGTTACATCCCT 788
OY 61 MetLeuGluThrArg-ArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAspPr 80
DB 789 ATGTTGGAACAAAGCAGGCGAGATTCTTCCTTATCTTAAACAGTAACAGCAATCC 848
OY 80 oLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
DB 849 TAAAGTCACTTTTAAAGATGTCAGAAATTAATGATTGATGCTGAAGAAAGATCATGATGT 908
OY 98 ----- 98
DB 909 CATGATTAAGAACTAATTCACCTTTTGAGATTCTATTGAAGAGCTGAATATATTTC 968
OY 98 ----- 98
DB 969 CAGAGCTGATGACTCCTGAGGAAATATAGTTGTCATCTACTAGATGCGTCCCA 1028
OY 98 ----- 98

```

```

Db      1029 GACTCGCTACAGATAGTGTGATCTCACCTGAATTATTATCCAGTAGAAGATGATGT 1088
QY      98 ----- 98
Db      1089 TATGAAAGGACAGAGCTCATTTAGTCACTGCTGATCTTGAGCCACATTAATTATCTA 1148
QY      98 ----- 98
Db      1149 TGAGAAACACACAGACATCTGATTAATAATCCATGACATCTTTCATGTTTTTCCCAAG 1208
QY      98 ----- 98
Db      1209 TCACGAGAGAAATTGAGTTATTTTGGCTCTGAAAGCAAAACAGTTCCGTCATTT 1268
QY      98 ----- 98
Db      1269 ATACAAAATTACCTATTATTTTAAAGAAACAAATATTAACGATCCAGTGGGCTGCC 1328
QY      98 ----- 98
Db      1329 TGCTCAACTGTCACTTGGATGATCAATTCATGAGATCTCTAGAACTCCATCCTGTAT 1388
QY      99 -----ValAspGluValArgArgLeuValTyrPheG 109
Db      1389 GTGTGTGACACATATAGTTGAGATCCAGTTGATGAAGTCAGAAAGCTGGTATATTTTGA 1448
QY      109 uGlyThrIlyAspSerProleuGluHisIleuTyrValIvalSerTyrValAsnProG 129
Db      1449 AGGACCAAAAGACTCCCTTTAGAGCATCACTGTACGATGATGATGATGATGATGATG 1508
QY      129 yGluValThrArgLeuThrAspArgIlyTyrSerHisSerCysCysIleSerGlnHisC 149
Db      1509 AGAGGTGACAGAGGCTGACTGACCGTGGCTACTCACAATTCCTTGTCATCAGTCAGACAG 1568
QY      149 sAspPhePheIleSerIlyTyrSerAsnGlnIlyAsnProHisCysValSerLeuTyrI 169
Db      1569 TGACTCTTTATTAAGTAAAGTATAGTAAACAGAAATTCACACTGTGTGCTTTTACAA 1628
QY      169 sleuSerSerProGluAspAspProThrCysIlyThrIlyGluPheTyrAlaThrIle 189
Db      1629 GCTATTAAGTCTCGAAGATGACCCCACTTGCAAAACAAAGAAATTTGGCCACCAATTT 1688
QY      189 uAspSerAlaGlyProleuProAspTyrThrProProGluIlePheSerPheGlnSerTh 209
Db      1689 GGAATCAGCAGAGCTCTTCTCTACTACTCTCTCAGAAATTTCTCTTTTGAAGTAC 1748
QY      209 rThrGlyPheThrLeuTyrGlyMetLeuTyrIlyProHisAspLeuGlnProGlyIlyS 229
Db      1749 TACTGATTTTACATTGTATGGGATGCTTCAACAGCCTCATGATCTACACCTGGAAAGAA 1808
QY      229 sTyrProThrValIleuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPh 249
Db      1809 ATATCTACTGCTGCTTCAATATATGTGTCTCTCAGGTGACCTTGTAATATATGTT 1868
QY      249 eIlyGlyValIlyTyrPheArgLeuAsnThrIleuAlaSerLeuGlyTyrValValVal 269
Db      1869 TAAAGAGTCAAGTATTTCCGCTTGATATCCCTAGCCTCTAGGTATATGTTGTATG 1928
QY      269 lIleAspAsnArgGlySerCysHisArgGlyLeuIlyAspGluGlyValAlaPheIlyTyr 289
Db      1929 GATAGCAACAGGGGATCTGTACCGAGGGCTTAATTTGAAGGGGCCCTTAAATATATA 1988
QY      289 sMetGlyGlnIleGluIleAspAspIleValGluGlyLeuGlnTyrLeuAlaSerArgTy 309
Db      1989 AATG----- 1992
QY      309 rAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyGlyTyrIleuSe 329
Db      1992 ----- 1992
QY      329 rIleuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProVa 349
Db      1993 -----GTTGCTATTGCTGGGGCCCGCACT 2015

```

```

QY      349 lThrLeuTyrPhePheTyrAspThrGlyTyrThr 360
Db      2016 CACTCTGTGATCTTCTATGATACAGATACAG 2049
RESULT 13
ID      ABR83332 standard; cDNA; 4685 BP.
XX      ABR83332;
AC      ABR83332;
DT      12-AUG-2002 (first entry)
DE      cDNA encoding human DPRP-1 splice variant #8.
XX      Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW      DPP4; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW      diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW      heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW      ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW      dyskinesia; reproductive disorder; inflammatory disorder;
KW      metabolic disorder; gene; ss.
XX      Homo sapiens.
XX      MO200231134-A2.
XX      18-APR-2002.
XX      12-OCT-2001; 2001WO-US31874.
XX      12-OCT-2000; 2000US-240117P.
XX      (FERR ) FERRING BV.
XX      Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX      WPI: 2002-444178/47.
XX      P-PSDB, oAB61601.
XX      DR      New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT      the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT      viral infections, cancers, allergies, neurological disorders, or pain
PT      -
XX      PS      Disclosure; Page 75-76; 113pp; English.
XX      CC      The present invention relates to the isolation of novel human serine
CC      proteases referred to as dipeptidyl peptidase IV (DPP4)-related
CC      and nucleic acids encoding them are useful for treating infections
CC      such as fungal, bacterial, protozoan and viral infections, particularly
CC      infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC      pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC      bulimia, Parkinson's disease, acute heart failure, hypotension,
CC      hypertension, urinary retention, osteoporosis, angina pectoris,
CC      stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC      psychotic and neurological disorders (e.g. anxiety, dementia, or
CC      schizophrenia), and dyskinesias. These may also be used in discovering
CC      therapeutic agents for the treatment of reproductive, inflammatory and
CC      metabolic disorders. ABR83332-ABR83343 encode human DPRP proteins.
XX
SQ      Sequence 4685 BP; 1430 A; 853 C; 991 G; 1411 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 2,26e-167 Length: 4685
Score: 1522.00 Matches: 312
Percent Similarity: 58.54% Conservative: 0
Best Local Similarity: 58.54% Mismatches: 1
Query Match: 78.74% Indels: 221
DB: 24 Gaps: 2

```

US-10-070-464-7 (1-360) x ABR83332 (1-4685)

QY 1 GIUGIUSPALAArgSerAlaGlyValAlaThrPheValLeuGlnGluIlePheSerPheGlu 20
 Db 943 GAAGAAGATGCCAGATCAAGCTGAGTGGCTACCTTGTCTCCAGAAAGAAATTTGATAGA 1002
 QY 21 TyrSerGlyTyrTrpTrpCysProIysAlaGluThrThrProSerGlyGlyIysIleLeu 40
 Db 1003 TATTCGGCTATGGGTGGTGCACAAAGCTGAACCACTCCAGGGTGGTAAATCTT 1062
 QY 41 ArgIleLeuTyrGluGlnAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
 Db 1063 AGAATTCATATAGAAAGAAATGATGATCTGAGGTGAATATTCATCTTACATCCCT 1122
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProIysThrGlyThrAlaAsnPro 80
 Db 1123 ATGTTGAAACACAGAGAGGGCAGATTCATCCGTATCTCTAAACAGGTACCAATCTT 1182
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
 Db 1183 AAGTCACTTTAAAGATGTCAGAAATATGATTTGATGCTGAAGAGATCATAGATGTC 1242
 QY 98 ----- 98
 Db 1243 ATAGATAGGAACATAATTCAACCTTTAGATTCTATTTGAAGAGATTGAATATATGCC 1302
 QY 98 ----- 98
 Db 1303 AGAGCTGATGAGCTCTGAGGAAATATGCTTGTCATCTACTATAGATGCTCCAG 1362
 QY 98 ----- 98
 Db 1363 ACTGCCCTACAGATAGTGTGATCTCACCTGAAATATTTATCCCACTAGAGATGATGTT 1422
 QY 98 ----- 98
 Db 1423 ATGAAAGGACAGACTCATTTGATGATGAGCTGATTCGTGAGCGCACTATATCTAT 1482
 QY 98 ----- 98
 Db 1483 GAAGAACAACAGACATCTGATTAATATCCATGACATCTTTCATGTTTCCCAAGT 1542
 QY 98 ----- 98
 Db 1543 CACGAGAGGAATTGAGTTTATTTTGGCTCTGATGCAAAACAGGTTCCGTATTTA 1602
 QY 98 ----- 98
 Db 1603 TACAAATATCATCTATTTTAAAGAAAGCAATATATAACATCCAGTGGTGGCTGCT 1662
 QY 98 ----- 98
 Db 1663 GCTCCAGTATTTCAAGTGCCTCTATCAAAAGAGATAGCAATTACAGTGGTGAATGG 1722
 QY 99 -----ValAspGluValArgArgLeuValTyr 107
 Db 1723 GAAGTCTTGGCGGCATGATCTATATCCAAAGTTGATGAGTGAAGGGTGGATATAT 1782
 QY 108 PheGluGlyThrLysAspSerProLeuGlnHisIleValValSerTyrValAsn 127
 Db 1783 TTGGAAGGACCAAAAGACTCCCTTTAGAGCATCACTGTAGTGTAGTACCTAAT 1842
 QY 128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSerGln 147
 Db 1843 CCTGAGAGGTGACAAAGCTGACTGACCGTGGCTACATCTTGTCTGACATCAGTCA 1902
 QY 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
 Db 1903 CACTGTGACTTCTTTATAGTATAGTATAGTAAACGAAGATCCACATGCTGTCCCTT 1962
 QY 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 187
 Db 1963 TACAAGCTATCAAGTCTGAAAGATGACCAACTTCAAAAACAAAGAAATTTTGGGCCAC 2022

QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
 Db 2023 ATTTCGATTCAGT----- 2036
 QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
 Db 2036 ----- 2036
 QY 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
 Db 2037 -----CCTCAGTGCAGTTGGGATATAT 2060
 QY 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerIleuGlyTyrVal 267
 Db 2061 CGGTTAAAGGAGTCAAGATTCGCTGATACCTTAGCTCTCTAGATTATGGTT 2120
 QY 268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGluValAspLys 287
 Db 2121 GTAGTGAATGACAAACAGGGGATCCGTGACCAAGGGCTTAAATTTGAAGGCCCTTTAAA 2180
 QY 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnThrLeuAlaSer 307
 Db 2181 TATTAATGGGTCAATATAGAAATTCATGATCAGGTGGAAGAGCTCCAAATATCTAGTTCT 2240
 QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyTyr 327
 Db 2241 CGATATGATTTTCATGACTAGTAGTGTGGGCATCCAGCGCTGGCTATAGAGATATAC 2300
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 Db 2301 CTCTCCCGATGGCAATTAATGACAGGTGCAAGTATCTTCAGGGTGTCTATGCTGGGGCC 2360
 QY 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
 Db 2361 CCAGTCACTCTGTGATCTTCTATGATACAGATACAGC 2399
 Db
 RESULT 14
 ABRK83331
 ID ABRK83331 standard; cDNA; 4676 BP.
 AC
 AC ABRK83331;
 AC
 DT 12-AUG-2002 (first entry)
 XX
 XX cDNA encoding human DPRP-1 splice variant #7.
 DE
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; interlility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinestia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 XX
 OS Homo sapiens.
 PN WO200231134-A2.
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 PR 12-OCT-2000; 2000US-240117P.
 PA (FERR) FERRING BV.
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 DR WPI; 2002-444178/47.
 DR P-PSDB; ABG61600.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 the proteins, useful for treating e.g. fungal, bacterial, protozoan and

PT viral infections, cancers, allergies, neurological disorders, or pain
PT -
XX
PS Disclosure, Page 72-73; 113pp; English.

CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl2 peptidase IV (DPP2V)-related
CC proteins (DPP2P). The dipeptidyl2 peptidase IV-related proteins (DPP2P)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis and/or pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABR83322-ABR83343 encode human DPP2P proteins.

SQ Sequence 4676 BP; 1424 A; 859 C; 979 G; 1414 T; 0 other;

Alignment Scores:

Pred. No.:	2,58=167	Length:	467
Score:	1521.50	Matches:	309
Percent Similarity:	57.97%	Conservative:	0
Best Local Similarity:	57.97%	Mismatches:	1
Query Match:	78.71%	Indels:	224
DB:	24	Gaps:	2

US-10-070-464-7 (1-360) X ABK83331 (1-4676)

QY	Glut1uaspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluInPheAspArg	20
Db	943 GAGAGAGATGCGAGATCACTGGAGTCCCTACCTTTGTTCTCCAGAAAGAAATTGATGA	1002
QY	21 TySerGlyIYrTPTpCySProLysAlaGluThrThrProSerGlyGlyLysIleLeu	40
Db	1003 TATTCTGGCTATTGGTGGTGTGCCAAACCTGAAACMACTCCACAGTGGTGAATTTCTT	1062
QY	41 ArgIleLeuYrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro	60
Db	1063 AGAATTCTATATGAAGAAATATGATCAATCTAGAGTGGAATATATTCATGTTACATCCCT	1122
QY	61 MetLeuGluThrArgArgAlaAspSerPheArgYrProLysThrGlyThrAlaAspPro	80
Db	1123 ATGTTGGAAACAAGAGGGCGAGATTCATTCGGTATCTTAACAAGGTACAGCAAACTCT	1182
QY	81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle-----	98
Db	1183 AAAGTCATTTTAAGATGTCAAGAAATATATGATTGATGTGAAGAGAGATCATGATGTC	1242
QY	98 -----	98
Db	1243 ATAGATAGAACTATATTCACAACCTTTTGAGATCTATTTGAAGAGGTGAATATATTGCC	1302
QY	98 -----	98
Db	1303 AGAGCTGATGSACTCTTGAGGGAAATATATGCTTGGTTCATCTTAATGATGCTCCAG	1362
QY	98 -----	98
Db	1363 ACTGGCCACAGATAGTGTGATGTCACTGAATTTATTATCCAGTAGAAGATGATGT	1422
QY	98 -----	98
Db	1423 ATGAAAAGCAGAGACTCATTTGAGTCAAGTGCCTGATTTCTGAGCGCACTAATTATCTAT	1482
QY	98 -----	98
Db	1483 GAAAGAACACAGACATCTGGATTAATATTCATGACATCTTTTCATGTTTTTCCCCAAGT	1542
QY	98 -----	98

Db	1543	CACGAAAGAGAAATTGAGTTATTTTGGCTTGAAATGCAAAACAGATTTCCGTCATTTA	1602
QY	98	-----	98
Db	1603	TACAAATATACATCTATTTTAAAGAAAGCAAAATATTAAGATCCAGTGGGTGCTCCT	1662
QY	98	-----	98
Db	1663	GCTCCAAAGTATTTCAAAGTCTCATCAAAAGAGATAGCAATTACAGTGGTAATGG	1722
QY	99	-----ValaspGluValAArgGLeuValTyr	107
Db	1723	GAAGTCTTGGCCGGCATGGATCTTAATATCCAACTATATGAAGTCAAGAGGCTGGATAT	1782
QY	108	PheGluGlyThrIlyAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn	127
Db	1783	TTTGAGGCACCAAAAGACTCCCTTTAGAGATCACTGTATCGTATAGTACGTAAAT	1842
QY	128	ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyLeuSerGln	147
Db	1843	CCTGAGAGGGGACAAAGCTGACGACCGGTGGCTACTACAACTCTTCTGATCAGCAG	1902
QY	148	HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu	167
Db	1903	CACGTGACTCTTTATATAGTATAGTATATGATACCAAGAAATCCACTGTGTCTCCTT	1962
QY	168	TyrIlyLeuSerSerProGluAspAspProThrCysLysThrIlyGluPheTyrPalaThr	187
Db	1963	TACAAGCATCAAGTCTCGAAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCCACC	2022
QY	188	IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu	207
Db	2023	ATTGTGATTCAGAGGCTCCTCTTCGACTATCTCCTCCAGAAATTTCTCTTTGAA	2082
QY	208	SerThrArgIlyPheThrIleuTyrGluMetLeuTyrIlyProHisAspLeuGlnProGly	227
Db	2083	AGTACTGATTTTACATGTGATGGATGGCTCTCAAGCCTCATGATCTACAGCCTCGGA	2142
QY	228	LysIlyTyrProThrValLeuPheIleTyrGlyGlyProGluValGlnLeuValAsnAsn	247
Db	2143	AAGAAATATCTACTAGTGTCTGTATATATGTGTGTG	2180
QY	248	ArgPheIlyGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal	267
Db	2180	-----	2180
QY	268	ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys	287
Db	2180	-----	2180
QY	288	TyrIlyMetGlyGlnIleGluIleAspAspGlnValGluIlyLeuGlnTyrLeuAlaSer	307
Db	2181	-----GGTCAAAATAGAAATATGACGATCAGTGGGAAGGACTCCAAATATCTAGCTTCT	2211
QY	308	ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyrGlyGlyTyr	327
Db	2232	CGAATATGATTCATGTGACTTAGATCGGTGGGATCCACCGCTGTGCTTATGGAGATAC	2291
QY	328	LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla	347
Db	2232	CTCTCCCGATGGCATTTATATCAGAGGTCACATATCTTCAGAGGTGGCTATTTGCTGGGGCC	2351
QY	348	ProValThrIleuThrIlePheTyrAspThrGlyTyrThr	360
Db	2352	CCAGTCACTCTGTGGATCTTCTATGATACAGGATACAG	2390
RESULT 15			
ABV76411			
ID	ABV76411	standard; cDNA; 1821 BP.	
XX	ABV76411;		
XX			

DT 07-FEB-2003 (first entry)

XX Dipeptidyl peptidase IV protein 27.5-encoding cDNA.

DE Dipeptidyl peptidase IV protein 27.5; recombinant production;

XX Stenotrophomonas maltophilia dipeptidyl peptidase IV homologue;

KW gene therapy; diabetes; tumour; cancer; cytostatic; gene; ss.

XX Unidentified.

OS

XX Key Location/Qualifiers

FT CDS 826..1578

FT /tag=a

FT /product="Dipeptidyl peptidase IV protein 27.5"

XX CN1342768-A.

XX 03-APR-2002.

XX 12-SEP-2000; 2000CN-0125127.

XX EF 12-SEP-2000; 2000CN-0125127.

XX PR 12-SEP-2000; 2000CN-0125127.

XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX PA

XX PI Mao Y, Xie Y;

XX DR WPI; 2002-529809/57.

XX DR P-PSDB; ABB99949.

XX

PT New dipeptidyl peptidase IV protein 27.5 and encoding polynucleotide,

PT useful for treating diabetes and tumors -

XX

XX Claim 6; Page 24-25 (Disclosure); 32pp; Chinese.

XX

CC The invention relates to dipeptidyl peptidase IV protein 27.5 (ABB99949)

CC and nucleic acids encoding it (ABV76411). The protein has a molecular

CC weight of 27.5 kD and has 40% identity and 58% homology over a 227 amino

CC acid stretch with Stenotrophomonas maltophilia dipeptidyl peptidase IV

CC (GenBank accession number D83263). The invention also relates to a method

CC for the recombinant production of the protein, an antagonist of the

CC protein, and the use of the protein, gene and antagonist in therapeutic

CC applications. Dipeptidyl peptidase IV protein 27.5 can be used in the

CC treatment of a variety of diseases such as diabetes and tumours. The

CC present sequence represents cDNA encoding dipeptidyl peptidase IV

CC protein 27.5.

XX

SQ Sequence 1821 BP; 533 A; 374 C; 391 G; 523 T; 0 other;

XX

Alignment Scores:

Pred. No.: 4.42e-157 Length: 1821

Score: 1429.00 Matches: 279

Percent Similarity: 84.84% Conservative: 12

Best Local Similarity: 81.34% Mismatches: 22

Query Match: 73.93% Indels: 30

DB: 24 Gaps: 6

US-10-070-464-7 (1-360) x ABV76411 (1-1821)

QY 43 LeuTYrGIuGIuAsnAspGIuSerGIuValGIuIleIle----- 55

DB 244 GTTTTCCCAAGAGCGAAGAGAAATTGAGTTATTTTGGCTCGAATGCAAAACA 303

QY 56 -----HisValThiSerProMetLeuGIuThiArgArgAlaaspSer 69

DB 304 GGTTCGTCATTTATACAAATTAACATCTATTTTAAAGAAAGCAAA----- 351

QY 70 PheArgTYrProLYsThiGIuThiAlaasnProLYsValThiPheLYs-----Met 86

DB 352 TATTAAGCATCAGTGGTGGCTGCTGCTCAAGT---GATTTCAAGTGTCTATCAAA 408

QY 87 SerGIuIleMetIle-----AspAlaGIuGIuArg-----Ile 97

DB 409 GAGGAGATAGCAATTACGATGTCGATAGGAAAGTCTTGGCCGCGCATGATCTATATC 468

QY 98 IleValAspGIuValArgArgLeuValTYrPheGIuGIuThiLYsAspSerProLeuGIu 117

DB 469 CAAGTGTATGAAGTCAGAAAGGCTGCTATATTTTGAAGCACACCPAAGACTCCCTTTAGAG 528

QY 118 HisIleuTYrValValSerTYrValAsnProGIuGIuValThiArgLeuThiAspArg 137

DB 529 CATCACCTGTAGCTAGTACGTTACGTAATCTCGAGAGGTGACAAAGCTGACTGACCCCT 588

QY 138 GIuTYrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLYsTYrSer 157

DB 589 GGCTACTACATCTTGGTCGATCAGTCAGACCTGTGACTTCTTATTAAGTAAGTATATG 648

QY 158 AsnGIuLYsAsnProHisCysValSerLeuTYrLYsLeuSerSerProGIuAspAspPro 177

DB 649 AACCAAGAAATCCACACATGTGTGTCCCTTTACAAAGCATATCAAGCTCGAAGATACCA 708

QY 178 ThCYsLYsThiLYsGIuPheTrpAlaThiIleuAspSerAlaGIuProLeuProAsp 197

DB 709 ACTTGCAAAACAAAGAAATTTGGGCCACCAATTTGGATTCAGCAGGTCTCTCTTGAC 768

QY 198 TYrThiProProGIuIlePheSerPheGIuSerThiThiGIuPheThiLeuTYrGIuMet 217

DB 769 TATCTCTCCAGAAATTTCTCTTTTAAAGTACTGATTTACATTTGATGAGATG 828

QY 218 LeuTYrLYsProHisAspLeuGlnProGIuLYsLYsTYrProThiValLeuPheIleTYr 237

DB 829 CTCTACAAAGCTCATGATCTACAGCTCGAAGAAATATCTCTACTGCTGCTGTCATATAT 888

QY 238 GIuGIuProGlnValGlnLeuValAsnAsnArgPheLYsGIuValLYsTYrPheArgLeu 257

DB 889 GGTGTCTCTCAGGTCAGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 948

QY 258 AsnThiLeuAlaSerLeuGIuTYrValValValIleAspAsnArgGIuSerCysHis 277

DB 949 AATACCTAGCCTCTCTAGGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1008

QY 278 ArgGIuLeuLYsPheGIuGIuValaPheLYsTYrLYsMetGIuGlnIleGIuIleAspAsp 297

DB 1009 CGAAGGCTTAATTTGAAGGCGCTTTAAATATAAATGGGTCAATGAAATGACCAT 1068

QY 298 GlnValGIuGIuLeuGlnTYrLeuAlaSerArgTYrAspPheIleAspLeuAspArgVal 317

DB 1069 CAGGTGGAAGGACCTCCATATATCTAGCTTCTGATATGATTTTCATTGACTTATGATGTGTG 1128

QY 318 GIuIleHisGIuTYrSerTYrGIuGIuTYrLeuSerLeuMetAlaLeuMetGlnArgSer 337

DB 1129 GGCAITCCAGGCTGCTCTATGAGAGATCTCTCTCGATGCGATTATATGACAGAGTCA 1188

QY 338 AspIlePheArgValAlaIleAlaGIuValaProValThiLeuTrpIlePheTYrAspThi 357

DB 1189 GATATCTTCAAGGTTGCTATGATGTGCGGCCCCAGTCACTGTGATCTTCTATATATCA 1248

QY 358 GIuTYrThi 360

DB 1249 GGATACACG 1257

Search completed: October 15, 2003, 17:58:24

Job time : 292.512 secs

Db 2486 GCGGTCGCGACTGATCTATGAGAGAGATCCTCGCGATCGTACGG--TTGTGG 2542
QY 26 TTPCSPProlyalAaIuThrThrProSerGlyGlyLeuAArgIleLeuTyrglu 45
Db 2543 TTCTGCGCGAGTGTAA-----TATCTGCTTACTTG 2575
QY 46 GUAAspGluSerGluValGluIleHisValThrSerProMetLeuGluThraag 65
Db 2576 AGCTTAATGAGACTGGGTTCCGACTACACCGCTCAGTATTATGATTAACAAGAG 2635
QY 66 ArgAlaAspSerPhe-----ArgTyProIysThrGlyThraIaasn 79
Db 2636 ATCGCTCCGCGCATCCATGGAGCTGAGATAGAGTATCCCAAGTGTGCGAGAGAA 2695
QY 80 ProIysValThr----- 83
Db 2696 CCGACCGTACGTTGAGTCTGCTTAATCATCGTAGCAAGAGTGAAGCAGCGCCGATC 2755
QY 84 -----PheIysMetSerGluIleMetIle----- 91
Db 2756 GACGCTTGAAGTCACTGACTGATTTGGGAGAGTTGCTGCTCACTGACTAC 2815
QY 91 ----- 91
Db 2816 ACCACCGTTGCTGCTAAGCGTTCAACCGTGTCCAGAGCCAGCAAAAGGTCTGCGGTC 2875
QY 92 -----AspAlaGluGlyArg----- 96
Db 2876 GATACGTCTCGAACAAGGCTACTGTCAITCAGCGCAGAGTGGACCCGATGATGGCTTC 2935
QY 96 ----- 96
Db 2936 GATACTTCTTCAATGAAGTATTGGCCCTATCAGCGCCGTCGACAAAGATGCTTAC 2995
QY 96 ----- 96
Db 2996 TACATGACATCTGTGACCATTCGGGATGGCGCATCTGATCTTCCCGTTTGGGCG 3055
QY 97 -----IleIleValasp 100
Db 3056 GCGGAACCTATCCCACTAACCAAGCGACTGGAGGTCACTTATTCAGTATTGAT 3115
QY 101 GluValArgArgIleuValTyrrPheGluGlyThrIysAspSerProLeuGluHisIleu 120
Db 3116 CAGGAACGCGAGTGGTGTACTACCTGTGACCTCAACACGACGACGCGCCATCTC 3175
QY 121 TyrValValSerTyrrValasnProGlyGluValThrArgLeuThraAspArgIlyTySer 140
Db 3176 TACTCGCTCTCTAT--TCCACGTTGGCGTACCCCGCTCTCGACACACCGTTGGC 3232
QY 141 -----HisSerCysIleSerGlnHisCysAspPheIleSerIysTyrrSerasn 158
Db 3233 GCGTACTGCTCTCTCTCTCTCGGAACTCGGGCTACTACATCTCATATCGAGGC 3292
QY 159 GlnIysAsnProHisCysValSerLeuTyrrIysLeuSerSerProGluAspProthr 178
Db 3293 CCAGAGCTACCCCTAC--CAGGAACCTTACACGACCAACAGT----- 3331
QY 179 CysIysThrIysGluTherIleThrIleLeuAspSerIleGlyProIeu----- 195
Db 3332 -----ACCAAAACCACTC--GCGACAACTACCGACAGCGCAAACTACTCGCAAAATC 3382
QY 196 ProAspTyrrThrProGluIlePheSerPheGlu-----SerThrArgIlyPheThr 213
Db 3383 AAGGACTATGCACTGCGCAACATCACCTTCTGACTTCCCTCCCTCCGAGAAAC 3442
QY 214 LeuTyrrGlyMetLeuTyrrIysProHisAspLeuGlnProGlyIysIysTyrrProThrVal 233
Db 3443 CTCATATGATGAGCGCTTACCCCGGGGTCTCCCGGATGAAGAGTAAAGTACCCATACTT 3502
QY 234 LeuPheIleTyrrGlyIysProGluValGlnLeuValAsnAspArgPheIysGlyValIys 253
Db 3503 TTCACCCCATACGCGCGGCCCAAGCGCCCAAGAAAGTACCAAGATGGCAGCCCTGAAT 3562

QY 254 Tyr---PheArgLeuAsnThrIleuAlaSerLeuGlyTyrrValValValIleAspAsn 272
Db 3563 TTCAAGCGCTATGTGCGCTCCGACAGCGAACTGAGTACCTGAGCTGTGCAAC 3622
QY 272 ArgGlySerCysHisArgGlyLeuIysPheGluGlyAlaPheIysTyrrIysMetGlyGln 292
Db 3623 CCGGACACAGGTTTCAAAAGACGCAAGTTCCGCTCCGCTACAGCGGCCAACTCGGCTTC 3682
QY 293 IleGluIleAspAspGlnValGluGlyLeuGlnTyrrLeuAlaSerArgTyrrAspPheIle 312
Db 3683 CTCGAAGCAGAAAGACCATCTACCGCGCAACAG---GCGGCAACATCCCTCGATC 3739
QY 313 AspLeuAspArgValGlyIleHisGlyTyrrPheSerTyrrGlyGlyTyrrIleuSerLeuMetAla 332
Db 3740 GATGACAGACACATCGCATCTGGGCTGTGAGTTCGAGGCTACTTGACAGCAAGTTC 3799
QY 333 LeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyValaProValThrLeuTyrr 352
Db 3800 CTGAGAAAGACACGCGTCTTTCACATTAAGAGTATCATCCGCTGTTCTGACTGG 3859
QY 353 IlePheTyrrAspThrGlyTyrrThr 360
Db 3860 CGTTCTACACTCAATGATACAG 3883

RESULT 2

US-09-280-116-171
; Sequence 171, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176365
; CURRENT APPLICATION NUMBER: US/09/280,116A
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 171
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: prolyl oligopeptidases
US-09-280-116-171

Alignment Scores:

	Pred. No.:	Length:
Score:	1,09e-31	823
Percent Similarity:	346.00	Matches: 67
Best Local Similarity:	98.57%	Conservative: 2
Query Match:	17,90%	Mismatches: 0
DB:	4	Indels: 1
		Gaps: 0

US-10-070-464-7 (1-360) x US-09-280-116-171 (1-823)

QY 292 GlnIleGluIleAspAspGlnValGluGly-LeuGlnTyrrLeuAlaSerArgTyrrAspPh 311
Db 3 CAATAGAAATGACATGAGTGTGAGAGACATCCATATCTAGCTTCTCGATATATTT 62
QY 311 eIleAspLeuAspArgValGlyIleHisGlyTyrrPheSerTyrrGlyGlyTyrrLeuSerLeu 331
Db 63 CATTGACTTAGATCGGTGGCATCCACCGCTGTCTTGTGAGAGTACTCTCCCTGAT 122
QY 331 fAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrIle 351
Db 123 GGCATTAAATGACAGGTCAGATATCTTCAGGGGTGTGCTATGTGCGGTCCCACTCACTCT 182
QY 351 vTrrIlePheTyrrAspThrGlyTyrrThr 360
Db 183 GTGATCTTCTATGATACGATACAG 210

RESULT 3
PCT-US93-07923-1

QY 308 ArgTYrAspPheIleAspLeuAspArgValIleHISglYTrpSerTYrGlyIYr 327
 Db 1853 AAAATGGAGATTGTGGACACAAACGATTCGACATTTGGGGCTGGTCATATGGAGGGTAC 1912
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaIleAla 347
 Db 1913 GTAACCTCAATGGTCTGGGATCAGGAAGTGGCGTTCCTCAAGTGTGAATAGCCCGGGG 1972
 QY 348 ProValThrLeuThrIlePheTYrAspTYrArgTYrThr 360
 Db 1973 CCTGATCCCGGTGGAGTACTATGACTAGCTAGTACACA 2011
 RESULT 4
 US-10-002-593-5
 / Sequence 5, Application US/10002593
 / Patent No. 6386198
 / GENERAL INFORMATION:
 / APPLICANT: Vanderbilt University
 / APPLICANT: Brown, Nancy J.
 / TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTIN
 / FILE REFERENCE: Acty Docker No. 6586198 1242/48/2
 / CURRENT FILING DATE: 2001-10-31
 / PRIOR APPLICATION NUMBER: US/10/002,593
 / PRIOR FILING DATE: 2000-10-31
 / NUMBER OF SEQ ID NOS: 10
 / SOFTWARE: PatentIn version 3.1
 / SEQ ID NO 5
 / LENGTH: 3407
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 US-10-002-593-5
 Alignment Scores:
 Pred. No.: 5,08e-27 Length: 3407
 Score: 315.00 Matches: 120
 Percent Similarity: 34.89% Conservative: 59
 Best Local Similarity: 23.39% Mismatches: 146
 Query Match: 16.30% Gaps: 188
 DB: 4 Gaps: 22
 US-10-070-464-7 (1-360) x US-10-002-593-5 (1-3407)
 QY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValIleuGlnGluGlu---Pheasp 19
 Db 643 AAAGAAGATATATATATATATGGAATGACTGGGTTTATGAGAGAAAGCTTTCAGT 702
 QY 20 ArgTYrSerGlyTYrTYrTYrPCysProIysAlaGluThrThrProSerGlyGlyIysIle 39
 Db 703 GCCTACTCTGCTCTGGTGGTGTCTCCAAACGGCAGCTTT----- 741
 QY 40 LeuArgIleLeuTYrGluGluAsnAspGlnSerGluValGluIleIleHis----- 56
 Db 742 -----TTAGCATATGCCCCAATTAAACGACAGAGATCCCACTTATGAAATATCTCTTC 795
 QY 57 ValThrSerProMetLeuGlnThrArgAlaAspSerPheArgTYrProIysThrGly 76
 Db 796 TACTCTGATGAGTCTGAGTACCAAGAGCTGACGGTTCATATCCAAAGCAGAGA 855
 QY 77 ThrAlaAsnProIysValThrPhe----- 84
 Db 856 GCTGTGAATCCCACTGTAAGTTCTTTGTGTAATACAGACTCTCTCAGCTCAGTCACACC 915
 QY 85 LysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAspGluValAlaArg 104
 Db 916 AATGCAACTTCATACAAATCACTGCTCTCTCATGTGATGAGG----- 963
 QY 105 LeuValTYrPheGluGlyThrIysAspSerProLeuGluHisIleuTYrValIysSer 124
 Db 964 -----GATCAGCTACTTGTGATGAGACA 987
 QY 125 TyrValAsnProGlyGluValThr-----ArgLeuThrAspArgGlyTYr 139

Db 988 TGGGGACACAGAAAGATTTCTTTCAGTGGCTCAGAGAGATTCAGAACTATTCGTC 1047
 QY 140 SerHisSerCys-----CysIle-----SerGln 147
 Db 1048 ATGATATATTGTGACTATGATGATATCCAGTGAGATGGAATGCTGTTAGTCACGGCAA 1107
 QY 148 HisCysAspPhe-----PheIleSerIysTYrSerAsnIlyAsnProHis 163
 Db 1108 CACATTGAAATGACTACTGCTGGCTGGTGAAGATTT--AGGCTTCAGAACTCAT 1164
 QY 164 CysVal-----SerLeuTYrIysLeuSerSerProGluAsp----- 175
 Db 1165 TTACCCCTTGATGTAATATGCTTCTTCAAGATCATCAGAAATGAGAGAGTTACACACAC 1224
 QY 176 -----AspProThrCysIysThrIysGluPheTYrAla 186
 Db 1225 ATTGCTATTTCCAAATAGATMAAAAGACTGCACATTTATTAAACAAAGCAGCTGGAA 1284
 QY 187 ThrIle----- 188
 Db 1285 GTCATCGGATAGAAAGCTCTTAACAGTATTATCTACTACTATTAATGAATATAA 1344
 QY 188 ----- 188
 Db 1345 GGATGCCAGAGAGAAAGATCTTATAAAATCCAATTATGACTATACAAAGTGACA 1404
 QY 188 ----- 188
 Db 1405 TGCCCTCAGTTGTAGCTGAATCCGAAAGTGTGACTATTCGTGTGTCATTCAGTAA 1464
 QY 189 -----LeuAspSerAlaGlyPro---LeuProAspTYrThr 199
 Db 1465 GAGCGAAGTATTATTCAGCTGAGATGTTCCGCTCTGCTGCTGCTCTTAATCTACAC 1524
 QY 199 ----- 199
 Db 1525 AGCAGCGTGAATGATMAAGGGCTGAGAGTCCGTAAGAACATTCAGCTTGTGATAAATG 1584
 QY 200 -----ProProGluIlePheSerPhe-----GluSerThrTYrGly 211
 Db 1585 CTGCAGAAATGCCAGATGCCCTCCAAAAAAGTGAATTTGATGATMAAACMAA 1644
 QY 212 PheThrLeuTYrGlyMetLeuTYrIysProHisAspLeuGlnProGlyIysTYrPro 231
 Db 1645 TTT---TGTATCAATGATCTTGGCTCTCAT---TTGATMAATCCAAAGAAATATCT 1698
 QY 232 ThrValLeuPheIleTYrGlyIysProGlnValGlnLeuValAsnAsnArgPheIysGly 251
 Db 1699 CTACTATTATGATGTATGACAGGCCCATGTACTCAA-----AAAGCA 1740
 QY 252 ValIysTYrPheArgLeuAsn-----ThrLeuAlaSerLeuGlyTYrValVal 268
 Db 1741 GACACTGTTTCAGCTGAACCTGAGCCACTTACCTTGCAAGACAGAAAACATTATAGTA 1800
 QY 269 Val---IleAspAsnArgGlySerCysHisArgGlyLeuIysPheGluIleAlaPheIys 287
 Db 1801 GCTACTTGTGATGAGAGAGAGTGTATACAGAGATATAGATCATGATCATCAAC 1860
 QY 288 TyrIysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTYrLeuAlaSer 307
 Db 1861 AGAAGACTGGGAACATTTGAAGTTGAAAGTCAAAATTTGAAGCCAGCAAAATTT--TCA 1917
 QY 308 ArgTYrAspPheIleAspLeuAspArgValGlyIleHISglYTrpSerTYrGlyIYr 327
 Db 1918 AAAATGGAGATTGTGGACACAAACGATTCGACATTTGGGGCTGGTCATATGGAGGGTAC 1977
 QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaIleAla 347
 Db 1978 GTAACCTCAATGGTCTGGGATCAGGAAGTGGCGTTCCTCAAGTGTGAATAGCCCGGGG 2037
 QY 348 ProValThrLeuThrIlePheTYrAspTYrArgTYrThr 360

Db 2038 COTGTATCCCGGTGGAGTACTATGACTCATGCTACACCA 2076

RESULT 5

US-09-221-017B-646/c

Sequence 646, Application US/09221017B

Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: ROSE, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Montoy, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 646:

SEQUENCE CHARACTERISTICS:

LENGTH: 657 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORIGINAL SOURCE:

ORGANISM: PORYPHYROMONAS GINGIVALIS

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1...657

US-09-221-017B-646

Alignment Scores:

Pred. No.: 2.45e-24 Length: 657

Score: 283.00 Matches: 60

Percent Similarity: 55.21% Conservative: 30

Best Local Similarity: 36.81% Mismatches: 69

Query Match: 14.64% Indels: 4

DB: 4 Gaps: 2

US-10-070-464-7 (1-360) x US-09-221-017B-646 (1-657)

QY 198 TyrThrProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMet 217

Db 566 TACAACCCGAAGAGATTTCACCACTATCAAACTCAATGGGGCTTGAACTGATGCTGG 507

QY 218 LeuTyrLysProHleAspLeuGlnProGlyLysTyrProThrValLeuPheIleTyr 237

Db 506 ATGCGAAGCTATTGATTTCGATCCCTCTCGCAGCTATCCCTGCTGATGTACAGTAT 447

QY 238 GlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeu 257

Db 446 AGCGGTCCCACTCCACAGAGTATGATGCTAT-----TCATTCATGGGANA 396

QY 258 AsnThrLeuAlaSerLeuGlyTyrValValVal11LeaPsnArgLysTyrCysHis 277

Db 395 CACTACCTTGATCGAAGAGTTTACCGTGGCATGTGTGATGGCGGCGACCGTGTCT 336

QY 278 ArgGlyLeuLysPheGluGlyValAlaPheLysTyrLysMetGlyGln11eGlu11LeaPsn 297

Db 335 CCGGCGAAGATGCGGCGAAGTGTACCTACATGCAACTGCGTGTATTGAAAGCGATGAT 276

QY 298 GlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPhe11LeaPsnAspArgVal 317

Db 275 CAGATAGCAGCGGCGCATGCTATAT--GAGCAGCTGCCCTATGTGATGATGACCTGATC 219

QY 318 GlyIleHisGlyTyrPserTyrGlyTyrLysSerLeuMetAlaLeuMetGlnArgSer 337

Db 218 GGCATATGGGGGTGAGCTATGGCGCTATACCACTAATGATGTTGTGTGGGGGAAAT 159

QY 338 AspIlePheArgValAla11eAlaGlyAlaProValThrLeuThr11PheTyrAspThr 357

Db 158 GGTACATTCAAGCGGGATAGCGCTTCTCTGTGGAGACTGCGGCTTTCTACCATTCG 99

QY 358 GlyTyrThr 360

Db 98 GTTACACC 90

RESULT 6

US-08-230-491A-1

Sequence 1, Application US/08230491A

Patent No. 5587299

GENERAL INFORMATION:

APPLICANT: Retzig, Wolfgang J.; Scanlan, Matthew J.;

APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.

TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR

TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES

TITLE OF INVENTION: THEREOF

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: FELFE & LYNCH

STREET: 805 THIRD AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: WORDPERFECT - ASC II

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/230,491A

FILING DATE: 20-APRIL-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 5587299man D.

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 330

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 688-3884

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2835 base pairs

TYPE: nucleic acid

```

; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-230-491A-1

```

Alignment Scores:	
Pred. No.:	3,05e-20
Score:	257.00
Percent Similarity:	36.70%
Best local Similarity:	23.51%
Query Match:	13,308
DB:	1
Length:	281
Matches:	114
Conservative:	64
Mismatches:	158
Indels:	150
Gaps:	20

US-10-070-464-7 (1-360) X US-08-230-491A-1 (1-2815)

Oy		8	GlyValAlaIatThrPheValLeuGlnGluIubHe-----AspArgYrSerGlyTyrTrp	25
Db		791	GGAAATCCCAAGACTGGGTATTATGAAGAAGAAATGCCTCCTACAAAATATGCT---CTCGG	847
Oy		26	TrpCySProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrlu	45
Db		848	TGGTCCTCT-----AATGGAAAAATTTTG-----GCATATACG	880
Oy		46	GluAsnAspGluSerGluValGluIleIleHisValThrSerPrometLeuGluThr---	64
Db		881	GAAATTATTAATGAATAGATATACCATGTTATTCGCTATTCCTATTATGGCGAACAATAT	940
Oy		65	ArgArgAlaAspSerPheArgTYrProLysThrGlyThrAlaAsnProLysValThrPhe	84
Db		941	CCTGGAAACAAATAAATATTCATACCACCAAAGGCTGAGCTAAGAAATCCCCTGTTCCGATA	1000
Oy		85	LysMet-----SerGluIleMetIleAspAla	93
Db		1001	TTTTATTATGATACCATCACCTTACCCTCGATAGTAGTCCCCAGAAAGTCCCTGTTCCACA	1060
Oy		94	-----GluGlyArgIle	97
Db		1061	ATGATAGCCCTCAAGATATTATTATTCAGTTGGCTCACGTGGGTACTGATGAACGAGTA	1120
Oy		98	IleValAspGluValArgArgLeuValTyrlueGu-----	109
Db		1121	TGTTTGAGTGGCTVAAAAGAGTCCGAAGTTCCTCGCTCATATGTGACTTCAG	1180
Oy		110	-----GlyThyLysAspSerPro---LeuGluHisIleLeuTyrlue-----	121
Db		1181	GAAGCTGGCAGACACTGGGATTGTCTCCAAGGCCAGGACATATGAAGAAGAAGCAACT	1240
Oy		122	-----ValValSerTyrlValAsnProGlylu	130
Db		1241	GGATGGCGTGGTGATCTTTGTGTTCAAGACCATTTTCAGTATGATGATTCGATTCGATC	1300
Oy		131	ValThrArgLeuThrAspArg---GlyTyrlSerHisSerCysIleSecGlnHisCys	149
Db		1301	TACAAATAATTTAGTACAAAGATGGCTACAAACATAT-TCACTATATCAANAAGACACTGT	1358
Oy		150	AspPhePheIleSerLysTyrlSerAsnGln-----	159
Db		1360	GSAAAATGCTATTCAAATTACAAAGTGGCAAGTGGAGGAGCCATAATATTATTCAGGTAAc	1419
Oy		160	-----LysAsnProHis	163
Db		1420	ACAGGATTCACGTGTTTATTCTTACGCAATGAATTTGAAGAAATACCTCGAAGAAAGAAAT	1479
Oy		164	CysValSer---LeuTyrlLysLeuSerSerProGluAsp-AspProThrCysLysThyly	182
Db		1480	CTACGAATTTAGCATTTGAGAGTATTCCTCCCAAGCAAGAAAGTGTACTTGGCATCTAAG	1533
Oy		182	SgLu-----PheTrpAlaThrIleLeuAspSerAla-----	192
Db		1540	GAAGAAGAGTGGCCATATTATACAGCAAGATTTCAGCATACGCAAGTACTATGACT	1599
Oy		193	-----GlyPro---LeuProAspTyrlThr-----	199
Db		1600	TGTCTCTAACGGCCCAAGGCAATCCCATTTTCCACCCCTTATGATGAGCAGCATGATCAAGA	1655

```

QY 200 -----Pr 200
Db 1660 AATTAATAATCTCGAGAGAAACAAGGATTTGAAAATGCTTTGAAAATATCCAGCTGCC 1719
QY 200 oProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLy 220
Db 1720 TAAAGAGAAATTAAGAACTGAGAGAGTAAAGTAAATTAATTAATTAATTAATTAATTA 1779
QY 220 sProHsaAspLeuGlnProGlyLysTyrProThrValLeuPheIleTyrGlyGlyPr 240
Db 1780 TCCTCTCAATTTGACAGATCAAAAGAGTATCCCTTGCTAATTCAGAGTATGATGGTGGTCC 1839
QY 240 cGlnValGlnLeuValAsnAsnAgnPheLysGlyValLysTyrPheArgLeuAsnThrLe 260
Db 1840 CTGACAGTCAGAGTGAAGTCT-----GTAATTCCTGTTAATTCGAT 1881
QY 260 uAlaSerLeu-----GlyTyrValValValValIleAspAsnArgGlySerCy 276
Db 1882 ATCTTATCTTCCAGTAGTAAAGAGGAGATGGTCATTCCTTGATGATGGTGAAGACAGC 1941
QY 276 sHisArgGlyLeuLysPheGluGlyIleAphelLysTyrLysMetGlyGlnIleGluLea 296
Db 1942 TTTCACAGGTGACAACTCCTCTATGACAGTATGCAAGAGCTGGGTGTTATGAAGTGA 2001
QY 296 pAspGluValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspAr 316
Db 2002 AGACCAAGTTACAGCTGTCAAGAAATTCAT--GAATGGGTTTCATTATGAAAAAG 2058
QY 316 gValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnAr 336
Db 2059 AATGACCATTAATGGGGCTGGTCTCTATGAGATAGATTTCATCACTGGCCCTTGATCTGG 2118
QY 336 gSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAs 356
Db 2119 AACTGGCTTTTCAAAATGGTATATAGACAGTGGCTCCAGCTCCAGCTGGGAATATTACG 2178
QY 356 pThrGlyTyrThr 360
Db 2179 GTCTGTCTACACA 2191

RESULT 7
; Sequence 1, Application US/08619280A
; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer, Park, John E.;
; APPLICANT: Reftig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; TITLE OF INVENTION: ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Feltz & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946

```

REFERENCE/DOCKET NUMBER: JUD 5330.1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 688-9200
 TELEFAX: (212) 838-3884
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2815 Base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 us-08-619-280A-1

Alignment Scores:
 Pred. No.: 3.05e-20 Length: 2815
 Score: 257.00 Matches: 114
 Percent Similarity: 36.70% Conservative: 64
 Best Local Similarity: 23.51% Mismatches: 158
 Query Match: 13.30% Indels: 150
 Gaps: 20

US-10-070-464-7 (1-360) x US-08-619-280A-1 (1-2815)

QY 8 GYValAlaThrPheValLeuGlnGluPhe-----AspArgTyrSerGlyTyrTrp 25
 Db 791 GGAATCCAGACTGGGTTTATGAGAGGAAATGCTTCCACAAATATGCT---CTCTGG 847
 QY 26 TTPCysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGlu 45
 Db 848 TGGTCTCCT-----AATGAAATTTTGG-----GCATATGCG 880
 QY 46 GIuaAspGluSerGluValGluIleIleHisValThrSerProMetLeuGlnThr--- 64
 Db 881 GAATTAAATGATGAAGATATACCAAGTATGCTTATCTCTATTATGCGCATGAACAAATAT 940
 QY 65 ArgAlaAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnProLysValThrPhe 84
 Db 941 CCTAAGACAAATTAATTCATCCATACCCAAAGCTGAGCTGAGATATCCCTTTCGAGTA 1000
 QY 85 LysMet-----SerGluIleMetIleAspAla 93
 Db 1001 TTTATTATGATACCACTTACCCCTGATGAGGTCCCCAGAAAGTGGCTTTCAGCA 1060
 QY 94 -----GluGlyArgIle 97
 Db 1061 ATGATAGCCTCAAGATATTATTATTCAGTGGCTCAGCTGGTTACTGATGAACAGATA 1120
 QY 98 IleValAspGluValArgGlyLeuValTyrPheGlu----- 109
 Db 1121 TGTTCAGTGGCTTAAAGAGTCCAGATGTTTGGTCTGTCTATATGTCATCTCAGG 1180
 QY 110 -----GlyThrLysAspSerPro--LeuGluHisLeuTyr----- 121
 Db 1181 GAAGACTGGCAGACATGGGATTTGTCAAAGACCCAGCATATGAGAAAGAACAGAACT 1240
 QY 122 -----ValIleSerTyrValAsnProGlyGlu 130
 Db 1241 GGATGGGCTGTGGATCTTTGTTCAAGACCAAGTTTCAGTTAAGATGCCATTTTCGATC 1300
 QY 131 ValThrArgLeuThrAspArg--GlyTyrSerHisSerCysCysIleSerGlnHisCys 149
 Db 1301 TACAAATATATTAGAGACAGAGTGTACAAACATAT-TGCTATATCAAAAGACACTGT 1359
 QY 150 AspPhePheIleSerLysTyrSerAsnGln----- 159
 Db 1360 GGAATATCTATTCAAATTTACAAAGTGCAAGTGGAAGGCCATTAATATTCAGATAAC 1419
 QY 160 -----LysAsnProHis 163
 Db 1420 ACAGATTCACGTGTTTATTTCTAGCAATGAATTTGAAGAAATACCTCGAAGAGAAACT 1479
 QY 164 CysValSer--LeuTyrLysLeuSerSerProGluAsp-AspProThrCysLysThrLys 182
 Db 1480 CTACAGATTTAGCATTTGGAAGCTATCTCCAAAGCAAGAAAGTGTTACTTGGCATCTAAG 1539

QY 182 sGlu-----PheTrpAlaThrIleLeuAspSerAla----- 192
 Db 1540 GAAAGAAAGTCCATATATACACAGCAAGTTCAGCGCACTACGCCAAGTACTATGCACT 1599
 QY 193 -----GlyPro--LeuProAspTyrThr----- 199
 Db 1600 TGTCTGTAAGGCGCCAGGCAATCCCATTTCCACCCCTTCATGATGAGACGCACTGATCAGA 1659
 QY 200 -----Pr 200
 Db 1660 AATTAAATCTCGAAGAAACAGAGAAATGGAATAATGCTTGAAAAATATCCAGCTGCC 1719
 QY 200 cProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
 Db 1720 TAAAGAGAAATTAAGAAACTTGAAGTAGATGAAATTAATCTTATGCTACAGATGATCT 1779
 QY 220 sProHisAspLeuGlnProGlyLysLysTyrProThrValIleLeuPheIleTyrGlyTyr 240
 Db 1780 TCCTCTCAATTGACAGATCAAGACAGATATCCCTTGTCTAAATTCAGATGATGTTGTTCC 1839
 QY 240 cGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLe 260
 Db 1840 CTGCGAGTCAGAGTGTAAAGTCT-----GATTTTCTCTTATATGAT 1881
 QY 260 uAlaSerLeu-----GlyTyrValValValAlaIleAspAsnArgLysCys 276
 Db 1882 ATCTTATCTTGCAATGAAGAGGATGTCATTCCTTCTGAGATGTCGAGAAACAGC 1941
 QY 276 sHisArgGlyLeuLysPheGluGlyValaPheLysTyrLysMetGlyGlnIleGluIleAs 296
 Db 1942 TTTCCAGGTGCAAACTCCCTTATGACGTATGAGAAAGCTGGGTTTATGAATTTGA 2001
 QY 296 pAspGlnValGluGlyLeuAsnGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 316
 Db 2002 AGACCAATATACAGCTGTACAGAAATTCATA---GAATGGCTTCATTCATGAGAAAG 2058
 QY 316 gValGlyIleHisGlyTyrPheTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 336
 Db 2059 AATAGCCATATGAGGCTGGCTCTATGAGAGATACGTTTCATCAGTGGCCCTTGCACTCGG 2118
 QY 336 sSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAs 356
 Db 2119 AACTGCTCTTTTCAATGTGTATAGCAGTGTCTCAGCTCCAGCTGCGAATATTACGC 2178
 QY 356 PThrGlyTyrThr 360
 Db 2179 GTCTGTCTACACA 2191

RESULT 8
 US-08-940-391-1
 ; Sequence 1, Application US/08940391
 ; Patent No. 5965373
 ; GENERAL INFORMATION:
 ; APPLICANT: Zimmermann, Rainer; Park, John E.;
 ; APPLICANT: Rectig, Wolfgang; Old, Lloyd J.
 ; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
 ; NUMBER OF INVENTIONS: PROTEIN ALPHA, AND USES THEREOF
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Felde & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/940,391

FILING DATE: 01-OCT-1997
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/619,280
 FILING DATE: 18-MARCH-1996
 APPLICATION NUMBER: 08/230,491
 FILING DATE: 20-APRIL-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Hanson, No. 5965373man D.
 REGISTRATION NUMBER: 30,946
 REFERENCE/DOCKET NUMBER: LUD 5330.1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 688-9200
 TELEFAX: (212) 838-3884
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2815 Base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-940-391-1

Alignment Scores:
 Pred. No.: 3,05e-20 Length: 2815
 Score: 257.00 Matches: 114
 Percent Similarity: 36.70% Conservative: 64
 Best Local Similarity: 23.51% Mismatches: 158
 Query Match: 13.30% Indels: 150
 Gaps: 20

US-10-070-464-7 (1-360) x US-08-940-391-1 (1-2815)

QY 8 G1ValAlaThrPheValLeuGlnGluPhe-----AspArgTyrSerGlyTyrTrp 25
 Db 791 GGAATCCAGCTGGGTTTATGAGAGAAATGCTTCACAAATATGCT--CTCTGG 847
 QY 26 TrpCysProLysAlaGluThrTrpSerGlyGlyLysIleLeuArgIleLeuTyrGlu 45
 Db 848 TGGTCTCT-----AATGAAATTTTGG-----GCATATGCG 880
 QY 46 GluAsnAspGluSerGluValGluIleIshIsvAlaThrSerProMetLeuGluThr--- 64
 Db 881 GAATTTATGATAGATATACAGATTATGCTATTCCTATATGCGCATGGAACATAT 940
 QY 65 ArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnProLysValThrPhe 84
 Db 941 CCTAGAACATTAATATTCATACCAAGGCTGAGCTAGATCCCGTTGTCGATA 1000
 QY 85 LysMet-----SerGluIleMetIleAspAla 93
 Db 1001 TTTATTTACATACCACTTACCTGCGTATGATAGTCCCGAGAGATGCTGTCAGCA 1060
 QY 94 -----GluGlyArgIle 97
 Db 1061 ATGATAGCTCAAGTATTTATTTATTCAGTTGCTCAAGGTTACTGATGATGAACGATA 1120
 QY 98 IleValAspGluValAlaGArgLeuValTyrPheGlu----- 109
 Db 1121 TGTTCAGTGGCTTAAAGAGTCACAGATGTTTCGCTCTGTATATGTGACTTCAGG 1180
 QY 110 -----GlyThrLysAspSerPro---LeuGluIshIsvLeuTyr----- 121
 Db 1181 GAAAGCTGGCAGATATGATGTTCCAAAGCCAGAGCATATAGAGAAAGAGAACT 1240
 QY 122 -----ValValSerTyrValAsnProGlyGlu 130
 Db 1241 GGATGGCTGTGATCTTTGTTTTCAGACCACTTATGATGATGATTCGATC 1300
 QY 131 ValThrArgLeuThrAspArg---GlyTyrSerHisSerCysGlyIleSerGlnHisCys 149
 Db 1301 TACAAATATTTATGATGACAGATGCTACCAACTAT-TCACTATATCAAGACACTGT 1359
 QY 150 AspPhePheIleSerLysTyrSerAsnGln----- 159

Db 1360 GGAATAATGCTATTCAAATTAACAAGTGGCAAGTGGAGGCCATTAATATTCAGATAC 1419
 QY 160 -----LysAsnProHis 163
 Db 1420 ACAGATTCAGTGTTTTATTTATTCAGCAATGAATTTGAAGATACCTCGAAGAAAGAACAT 1479
 QY 164 CysValSer---LeuTyrLysLeuSerSerProGluAsp-AspProThrCysLysThrIly 182
 Db 1480 CTACAGAAATTAAGATTGAAGGATTCCTCCAAAGCAAGAGTGTATTCCTGATCTTAAG 1539
 QY 182 sglu-----PheThrAlaThrIleLeuAspSerAla----- 192
 Db 1540 GAAAGAAAGGTGCCAATATTACACAGCAAGTTTCAGCGACTACGCCAAGTACTATGCAT 1599
 QY 193 -----GlyPro---LeuProAspTyrThr----- 199
 Db 1600 TGTCTGTACGGCCAGGATCCCATTTCCACCTTCATGATGACGACATCAAGA 1659
 QY 200 -----Pr 200
 Db 1660 AATTAAATCCTGAGAAAGAAACAGAAATTGAAATGCTTTGAAATATTCAGCTGCC 1719
 QY 200 oProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIly 220
 Db 1720 TAAAGGAAATTAAGAACTTGAAGTATGAAATTAATCTTATGATCAAGATGATCT 1779
 QY 220 sProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyIlyPr 240
 Db 1780 TCCTCTCATATTGACAGATCAAGAAAGTATCCCTGCTAATTCAGATGATGTGTCTCC 1839
 QY 240 oGlnValGlnLeuValAsnAsnArgPheLysGlyValIlyTyrPheArgLeuAsnThrIle 260
 Db 1840 CTCAGCTCAAGTGAAGTGGTCT-----GATTTCTGTAAATGAT 1881
 QY 260 uAlaSerLeu-----GlyTyrValIleValIleAspAsnArgIlySerCys 276
 Db 1882 ATCTTATCTTGGCAAGTAAAGAGGATGCTATTCCTGTGATGCTGAGGAGAACAGC 1941
 QY 276 sHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGlyLeas 296
 Db 1942 TTTCCAAAGTGACAAATCTCTTATGACAGTATCGAAAGCTGGGTGTTATGAAAGTTGA 2001
 QY 296 pAspGlnValGluGlyLeuGlnIlyLeuAlaSerArgTyrAspPheIleAspLeuAspAr 316
 Db 2002 AGACAGATTAACAGCTGTCAAGAAATTCATA---GAAATGGTTCTATGATGAAAAAG 2058
 QY 316 gValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnAr 336
 Db 2059 AATAGCATATGGGGCTGCTGCTATGAGAGATACGTTTCATCACTGGCCCTTGATCTGG 2118
 QY 336 gSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAs 356
 Db 2119 AACTGGCTTTTCAAAATGTGTATAGCAGTGGCTCCAGTCCAGCTGGGAATTTACGC 2178
 QY 356 pThrGlyTyrThr 360
 Db 2179 GTCTGTCTACACA 2191

RESULT 9

US-09-221-017B-253
 Sequence 253, Application us/09221017B

Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

```

; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221.017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P11182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P11546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P12811
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 253:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...543
; US-09-221-017B-253

Alignment Scores:
Pred. No.: 6,286-19 Length: 543
Score: 236.50 Matches: 41
Percent Similarity: 65.00% Conservative: 24
Best Local Similarity: 41.00% Mismatches: 34
Query Match: 12.23% Indels: 1
DB: 4 Gaps: 1

US-10-070-464-7 (1-360) x US-09-221-017B-253 (1-543)
QY 260 LeuAlaSerLeuGlnTYrValValValAlaLeuAspAsnArgGlySerCysHisArgGly 279
DB 43 ATGCACAGAAAGGCTATGCGCTTACGCGATGATCGCGGATCGCCATAGAGGG 102
QY 280 LeuIysPheGlnGlyAlaPheLysTYrLysMetGlyGlnIleGlnIleAspAsnVal 299
DB 103 GCTGCTTCGAGCAGGTTATTCATGCTGCTTGGGCGAGACGAGATGCGCGATCAGATG 162
QY 300 GlnGlyLeuGlnTYrLeuAlaSerArgTYrAspPheIleAspLeuAspArgValGlyIle 319
DB 163 TCGCGTGTGATTTCTCTCAAGAGCCAA---TCATGCGTGTGATTCGCGATGAGATG 219
QY 320 HisGlyTPSerTYrGlyGlyTYrLysSerLeuMetAlaLeuMetGlnArgSerAspIle 339
DB 220 CATGGCTGAGAGTATGATGCTTATGACTAGATCTGATGCTTACGACGCGCATGTC 279
QY 340 PheArgValAlaIleAlaGlyAlaIleProValThrLeuTrpIlePheTYrAspPheTrpLys 359
```

```

DB 280 TTCAAGTCGAGTACCGGCGGCTGTCTCACTAGACTGGAATTCGATATGAGATTATGTAC 339
RESULT 10
US-09-389-681-428
; Sequence 428: Application US/09389681A
; Patent No. 6518237
; GENERAL INFORMATION:
; APPLICANT: Yugu, Jiang
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.470C3
; CURRENT APPLICATION NUMBER: US/09/389,681A
; CURRENT FILING DATE: 1999-09-02
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 428
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-389-681-428

Alignment Scores:
Pred. No.: 1,4e-18 Length: 535
Score: 233.50 Matches: 52
Percent Similarity: 54.86% Conservative: 27
Best Local Similarity: 36.11% Mismatches: 54
Query Match: 12.08% Indels: 11
DB: 4 Gaps: 3

US-10-070-464-7 (1-360) x US-09-389-681-428 (1-535)
QY 221 ProHisAspLeuGlnProGlyLysLysTYrProThrValLeuPheIleTYrGlyPro 240
DB 15 COTCCTCAATTGACAGATCAAGAGATGATCCCTTGCAATTCAGATGATGATGCTCC 74
QY 241 GlnValGlnLeuValAsnAsnArgPheLysGlyValLysTYrPheArgLeuAsnThrLeu 260
DB 75 TGCAGTCAGAGTGAAGTCT-----GTTTGGCTGTTAATGGATA 116
QY 261 AlaSerLeu-----GlyTYrValValValAlaLeuAspAsnArgGlySerCys 276
DB 117 TCTTATCTTGCAGATGAGAGGAGTGTCTTGTGCTGATGATGTCGAGAACAGCT 176
QY 277 HisArgGlyLeuLysPheGlnGlyAlaPheLysTYrLysMetGlyGlnIleGlnIleAsp 296
DB 177 TTCAGAGGTGACAACTCTCTATGCACTGATGCAAGAGCTGGGTTTATGAGATTGAA 236
QY 297 AspGlnValGlnGlyLeuGlnTYrLeuAlaSerArgTYrAspPheIleAspLeuAspArg 316
DB 237 GACCGATTTACAGCTGTGAGAAATTCATA---GAAATGGCTTTCATGATGAAAAAGA 293
QY 317 ValGlyIleHisGlyTYrTPSerTYrGlyGlyTYrLysSerLeuMetAlaLeuMetGlnArg 336
DB 294 ATAGCCATATGGGGCTGTCTCTATGAGAGATACGTTTCATCATCGCCCTTCATCTGGA 353
QY 337 SerAspIlePheArgValAlaIleAlaGlyAlaIleProValThrLeuTrpIlePheTYrAsp 356
DB 354 ACTGCTTTTCAAAATGTGTATAGCAGTGTCTCCAGCTCGAGGAATATTATACGG 413
QY 357 ThrGlyTYrThr 360
DB 414 TCTGTCTACACA 425
RESULT 11
US-09-620-405B-428
; Sequence 428: Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
```



```
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/229,059
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,103
FILING DATE: 16-AUG-1996
APPLICATION NUMBER: 60/002,581
FILING DATE: 17-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 09272/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/322-5070
TELEFAX: 650/854-0875
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4982 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-229-059-1

Alignment Scores:
Pred. No.: 2.26e-14 Length: 4982
Score: 211.00 Matches: 49
Percent Similarity: 50.74% Conservative: 20
Best Local Similarity: 36.03% Mismatches: 55
Query Match: 10.92% Indels: 12
DB: 4 Gaps: 4

US-10-070-464-7 (1-360) x US-09-229-059-1 (1-4982)
QY 230 TyrProthrValleuPheileTYrGLyGLyProGlnValGlnLeuValIasnAsnArgPhe 249
Db 4941 TATCCTGATTTTCTTTCATATGSGGACCGAATTCACACAGTT-----4894
QY 250 LysGlyValIysTYrPheArgLeuAsn-----ThrLeuAlaSerLeuGlyTYr 265
Db 4893 -----GTCAAAAGCTTTCCGTAGCATTTAATGAGTGTAGCTTCACAAATTAAACGA 4840
QY 266 ValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlnGlyAla 285
Db 4839 ATGTGATGTGTGTGACGGTCTGTGACTGGCTTCAAGGTCAAGACTTTAGATCCCTT 4780
QY 286 PheLysTYrLysMetGlyGlnIleGluIleAspAspGlnValGlnGlyLeuGlnTYrLeu 305
Db 4779 GTTCGCATAGGCTCGGTGATTAACAGGCCCGCACCAAAATCTGCGGCTTCCTTATAT 4720
QY 306 AlaSerArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTTPSerTYrGly 325
Db 4719 GGTTCCT--TTAACTTTGTGTGATCCCAAAAGATTTCCTATTGTTGTGTGATACGGG 4663
QY 326 GlyTYrLeuSerLeuMetAlaLeuMetGlnArgSerAsp--IlePheArgValAlaIle 344
Db 4662 GGGTACCTGACACTTAACTTGGAGAAAGATCGCGAAGACATTTCAAAATTACGGGATG 4603
QY 345 AlaGlyAlaProValThrLeuTrpIlePheTYrAspThrGlyTYrThr 360
Db 4602 TCACTTGCCGACGATACCGACTGAGATTATACGATTCTGTATACT 4555
```

Search completed: October 16, 2003, 03:31:25
Job time : 81.4006 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 16, 2003, 00:05:32 ; Search time 269.152 Seconds
(without alignments)
3512.533 Million cell updates/sec

Title: US-10-070-464-7

Perfect score: 1993
Sequence: 1 EEDARSAGVATFTVQEEFDR.....RVATAGAPVTIMTYDTCYT 360

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1750203 seqs, 1313063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODL=ftime+ p2n.model -DEV=x1h
-O=/cgn2_1/USPTO.spool/US10070464/runat_15102003_113555_24924/app_query.fasta_1.2652
-DB=Published Applications NA -OPMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOBL=0 -LOOBLT=0 -UNITS=bits -SMAP=1 -END=1 -MATRIX=blcsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pct -NORM=ext -HEPSTIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10070464@cgn 1.1.845 @runat_15102003_113555_24924
-NCPU=6 -ICPU=3 -NO MMAP -LARGESUBSTR -NEG_SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAOP=6 -FGAEXT=7 -YGAOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PC1_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PC105_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match length	DB ID	Description
1	1880	97.3	4309	10 US-09-976-674-14 Sequence 14, Appl

2	1836.5	95.0	2649	12 US-10-054-776-1	Sequence 1, Appl
3	1836.5	95.0	2649	12 US-10-170-789-39	Sequence 39, Appl
4	1836.5	95.0	2671	10 US-09-976-674-2	Sequence 2, Appl
5	1836.5	95.0	3143	12 US-10-170-789-37	Sequence 37, Appl
6	1816.5	94.0	4829	10 US-09-976-674-12	Sequence 12, Appl
7	1522	78.7	4685	10 US-09-976-674-22	Sequence 22, Appl
8	1522.5	78.7	4676	10 US-09-976-674-20	Sequence 20, Appl
9	1229	63.6	4523	10 US-09-976-674-8	Sequence 8, Appl
10	1220.5	63.1	2617	10 US-09-976-674-4	Sequence 4, Appl
11	1220.5	63.1	4329	10 US-09-976-674-28	Sequence 28, Appl
12	1220.5	63.1	4302	10 US-09-976-674-24	Sequence 24, Appl
13	1201	62.1	4180	10 US-09-976-674-36	Sequence 36, Appl
14	1201	62.1	4263	10 US-09-976-674-34	Sequence 34, Appl
15	1183.5	61.2	3262	13 US-10-098-841-83	Sequence 83, Appl
16	1169.5	60.5	2801	13 US-10-098-841-100	Sequence 100, Appl
17	1132.5	58.6	4076	10 US-09-976-674-32	Sequence 32, Appl
18	1132.5	58.6	4159	10 US-09-976-674-30	Sequence 30, Appl
19	1113	57.6	4037	10 US-09-976-674-40	Sequence 40, Appl
20	1113	57.6	4120	10 US-09-976-674-38	Sequence 38, Appl
21	630	32.6	502	11 US-09-918-995-19585	Sequence 19585, A
22	368	19.0	561	11 US-09-764-891-877	Sequence 877, A
23	352	18.2	2079	13 US-10-044-090-843	Sequence 843, Appl
24	352	18.2	2411	10 US-09-976-674-26	Sequence 26, Appl
25	315	16.3	3407	12 US-10-423-714-5	Sequence 5, Appl
26	315	16.3	3407	13 US-10-002-593-5	Sequence 5, Appl
27	315	16.3	3407	14 US-10-165-603-6	Sequence 6, Appl
28	308.5	16.0	2130	10 US-10-156-761-1311	Sequence 3131, Appl
29	308.5	16.0	9025608	14 US-10-156-761-1	Sequence 1, Appl
30	297.5	15.4	4835	10 US-09-917-800A-1570	Sequence 1570, Appl
31	297.5	15.4	4835	10 US-10-165-603-5	Sequence 5, Appl
32	271	14.0	4797	11 US-09-764-891-7074	Sequence 7074, Appl
33	261	13.5	2812	14 US-10-084-817-24	Sequence 24, Appl
34	258.5	13.4	2366	10 US-10-101-510-683	Sequence 683, Appl
35	258.5	13.4	2788	12 US-10-269-009-22	Sequence 22, Appl
36	258.5	13.4	3138	14 US-10-198-846-13171	Sequence 13171, A
37	257.5	13.3	3224	12 US-10-240-965-117	Sequence 117, Appl
38	257	13.3	2814	10 US-09-962-832-108	Sequence 108, Appl
39	257	13.3	2814	10 US-09-954-456-1148	Sequence 1148, Appl
40	257	13.3	2814	10 US-09-954-531-367	Sequence 367, Appl
41	257	13.3	2814	12 US-10-101-510-30	Sequence 30, Appl
42	257	13.3	2814	12 US-10-301-822-54	Sequence 54, Appl
43	257	13.3	2814	12 US-09-873-367C-499	Sequence 499, Appl
44	257	13.3	2814	14 US-10-177-293-135	Sequence 135, Appl
45	257	13.3	2815	9 US-09-265-606-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-976-674-14
; Sequence 14, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akimsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Julien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 4309
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-14
Alignment Scores: 5.23e-231 Length: 4309
Pred. No.: 1

Score: 1880.00 Matches: 358
Percent Similarity: 98.90% Conservative: 0
Best Local Similarity: 98.90% Mismatches: 2
Query Match: 97.26% Indels: 3
DB: 10 Gaps: 0

US-10-070-464-7 (1-360) x US-09-976-674-14 (1-4309)

QY 1 GIUGIUAAPALAArSeraIaGlyValAlaThrPheValLeuGluGluPheAspArg 20
DB 943 GAAGAAGATGCGACGATCAGCTGAGTCCGTACCTTGTCTCCAGAGAAATTTATAGA 1002
QY 21 TTrSerGIYTrTTrPTrCysProLysaIaGluTrThrProSerGIYGLYysIleu 40
DB 1003 TATTCGGCTATTGGTGGTGGTCCAAAGCTCCCAAGCTCCAGTGGTGGTAAATTCCT 1062
QY 41 ArgIleuYrGIUGluAsnAspGluSerGIUValGluIleIshIsvAlThrSerPro 60
DB 1063 AGAATTCATATGAAGAAATGATGATCTGAGTGGAAATTATTCATGTTACATCCCT 1122
QY 61 MetLeuGIUThrArgArgAlaAspSerPheArgIYrProLysThrGIYThrAlaAsnPro 80
DB 1123 ATGTTGGAAACAAGAGGCGACATTCATTCGTTATCCTAACAAGTACAGCAATCCT 1182
QY 81 LysValIThrPheLysMetSerGIUleuMetIleAspAlaGluGIYArgIleIleValAsp 100
DB 1183 AAAGTCACCTTTAAAGATGTCAGAAATTAATGATGATGCGAAGGAATCCAA-GTTGAT 1241
QY 101 GIUValArgArgLeuValITrPheGIUGlyIThrLysAspSerProLeuGluIshIshIleu 120
DB 1242 GAAGTCAGAGGCTGGTATATTTGAAGGACCAAGATCCCTTATAGATCATCACCTG 1301
QY 121 TyrValIAserITyrValAsnProGIYGIUValIThrArgLeuThrAspArgIYTrSer 140
DB 1302 TACGATGATGATGATGATATCTGAGAGGTGACAGGCTACGACCGTGCTACCTCA 1361
QY 141 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysITrSerAsnGluLys 160
DB 1362 CATTCCTGGTCATCAGTCAGACAGTGTGCTCTTATTAAGATAGATAGTAAACAGAA 1421
QY 161 AsnProHisCysValIserLeuITrLysLeuSerSerProGluAspAspProThrCysLys 180
DB 1422 AATCCACACTGTGTGCCCTTTACCAAGCTATCAAGTCCCGAAGATGACCCCACTTGCAA 1481
QY 181 ThrLysGluPheITrAlaThrIleLeuAspSerAlaGlyProLeuProAspITrThrPro 200
DB 1482 ACAAGAGAAATTTGGGCCACATTTTGAITGACAGAGTCCCTTCCTGACTACACTCCT 1541
QY 201 ProGluIlePheSerPheGluSerThrITrGIYpHeITrLeuITrGIYMetLeuITrLys 220
DB 1542 CCAGAAATTTCTCTTTGAAAGTACTACTGATTTACATGTATGGGATGCTCTACAAAG 1601
QY 221 ProHisAspLeuGlnProGIYLysLysITrProITrValLeuPheIleITrGIYGLY-P 240
DB 1602 CCTCATGATCTACAGCTGGAAGAAATATCTACTAGTCTGTCTATATATGATGATCTC 1661
QY 240 ProGluValGluLeuValIAsnAspPheLysGIYValLysITrPheArgLeuAsnThrL 260
DB 1662 CTCAGGTGCAATGGTGGATTAATCGGTTAAAGGATCAAGTATTTCCGCTTGAAATCCC 1721
QY 260 euAlIserLeuGIYTrValIValIValIleAspAsnArgLysSerCysShIAsrGIYL 280
DB 1722 TAGCCTCTCTAGATTATGTGTGTGTAGTATAGACAAACAGGGATCTCTGACCAAGG 1781
QY 280 euLysPheGluGluValIAspLysITrLysMetGIYGlnIleGluIleAspAspGIYValG 300
DB 1782 TTAAATTTGAAGCGCTTTAAATATAAATAGAGATGCGATAGAGGTG 1841
QY 300 IUGIYLeuGluITrLeuAlaIserArgITrAspPheIleAspLeuAspArgValIGlyIleH 320
DB 1842 AAGGACTCCAATATCTAGCTTCTCGATATGATTTGATGACTTAGATGCTGAGGGATCC 1901
QY 320 IsgITrPserITrGIYGIYTrLeuSerLeuMetAlaLeuMetGlnITrGserAspIleP 340

DB 1902 ACGGTGATCTTATGAGATACCTCTCCGTATGATGATTAAGAGGTAGATATCT 1961
QY 340 hearValAlaITleAlaGlyAlaProValIThrLeuITrIlePheITrAspITrGIYTr 360
DB 1962 TCAGGGTGTATTCTCTGGGGCCAGTCACCTCTGTGATCTTATATATACAGATACA 2021

QY 360 hr 360
DB 2022 CG 2023

RESULT 2

US-10-054-776-1
; Sequence 1, Application US/10054776
; Publication No. US20030163618A1
; GENERAL INFORMATION:
; APPLICANT: Mark Robert Edbrooke
; APPLICANT: Alan Peter Lewis
; TITLE OF INVENTION: NOVEL PROTEIN
; FILE REFERENCE: OGI042US
; CURRENT APPLICATION NUMBER: US/10/054,776
; CURRENT FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2649)
US-10-054-776-1

Alignment Scores:

Pred. No.: 9,81e-226 Length: 2649
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 12 Gaps: 1

US-10-070-464-7 (1-360) x US-10-054-776-1 (1-2649)

QY 1 GIUGIUAAPALAArSeraIaGlyValAlaThrPheValLeuGluGluPheAspArg 20
DB 730 GAAGAAGATGCGACGATCAGCTGAGTCCGTACCTTGTCTCCAGAGAAATTTATAGA 789
QY 21 TTrSerGIYTrTTrPTrCysProLysaIaGluTrThrProSerGIYGLYysIleu 40
DB 790 TATTCGGCTATTGGTGGTGGTCCAAAGCTGAAACAACTCCAGTGGTGGTAAATTCCT 849
QY 41 ArgIleuYrGIUGluAsnAspGluSerGIUValGluIleIshIsvAlThrSerPro 60
DB 850 AGAATTCATATGAAGAAATGATGATCTGAGGAGAAATTAATTCATGTTACATCCCT 909
QY 61 MetLeuGIUThrArgArgAlaAspSerPheArgIYrProLysThrGIYThrAlaAsnPro 80
DB 910 ATGTTGGAAACAAGAGGCGACATTCATTCGTTATCCTAACAAGTACAGCAATCCT 969
QY 81 LysValIThrPheLysMetSerGIUleuMetIleAspAlaGluGIYArgIleIle- 98
DB 970 AAAGTCACCTTTAAAGATGTCAGAAATTAATGATGATGCTGAAAGAGATCATAGATGTC 1029
QY 98 ----- 98
DB 1030 ATAGATTAAGAACTATATTCACCTTTTGAATTCATTTTGAAGAGCTGAATATATGTC 1089
QY 98 ----- 98
DB 1090 AAGCTGATGATGATCTCTGAGGAAATATATGTTGCTCATCTACTAGATGCTCCAG 1149
QY 98 ----- 98

Db 1150 ACTGCGCTGAGATAGTGTGATCTCACCTGAATTATTATCCAGTAGAAGATGATGTT 1209
QY 98 ----- 98
Db 1210 ATGGAAGGACAGACTCATTTGAGTCAAGTCCGTATCTGTGACGCCACTAATTATCTAT 1269
QY 98 ----- 98
Db 1270 GAAGAACACACAGACATCTGGATTAATATCCATGACATCTTTTCATGTTTTTCCCAAGT 1329
QY 98 ----- 98
Db 1330 CACGAGAGGAATTGAGTTTATTTTGGCTCTGATGCAAAACAGTTTCCGTCATTTA 1389
QY 98 ----- 98
Db 1390 TACAATAATTACATCTATTTTAAAGGAACAATAATAACGATCCAGTGTGGCTGCT 1449
QY 98 ----- 98
Db 1450 GCTCCAGTATTTCAAGTGTCTTATCAAGAAGGAGATGCAATTACAGTGTGAAATGG 1509
QY 99 -----ValAspGluValArgLeuValTyr 1510
1510 GAAGTCTTGCGCGGCATGATCTAATATCCAAAGTTGATGAAGTCAGAAAGCTGGTATAT 1569
QY 108 PheGluGlyThrIleAspSerProLeuGluHisIleuTyrValValSerTyrValAsn 1629
1570 TTTGGAAGGACCAAGACTCCCTTTAGAGCATCACTGTCAGTGTGAGTCACTTAAT 1629
QY 128 ProGluGluValThrArgLeuThrAspArgIleTyrSerHisSerCysCysIleSerGln 147
1630 CCGTAGAGGTGACAGAGGCTGACCTGACCGTGGCTACATCTTCTGACATCCAGTCA 1689
QY 148 HisCysAspPhePheIleSerIleTyrSerAsnGlnYasnProHisCysValSerLeu 167
1690 CACGTGACTCTTATATAGTATAGTATAGTAAACGAAAGATCCACACGTGTGTCCCTT 1749
Db 168 TyrIleuSerSerProGluAspAspProThrCysIleTyrIleGluPheTrpAlaThr 187
1750 TACAAGCTATCAAGCTCCTGAAGATGACCACTTCCAAACAAAGAAATTTTGGGCCACC 1809
QY 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207
1810 ATTTTGGATTCAAGAGTCTCTCTTCTGACTATCTCTCCGAAATTTTCTCTTTGAA 1869
QY 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleProHisAspLeuGlnProGly 227
1870 AGTACTACTGAGTTTACATTTGATGATGATGCTCTCAAGCCCTCATGATCTACAGCTGGA 1929
Db 228 LysIleTyrProThrValLeuPheIleTyrGlyIleProGluIleGluLeuValAsnAsn 247
1930 AAGAAATATCTTACTGCTCTGCTCATATATGCTGCTCAGGTGCAGTGTGTAATAT 1989
QY 248 ArgPheIleGlyValIleTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
1990 CGGTTTAAAGGAGTAAAGTATTTCCGCTTGAATACCTTACGCTCTCTGAGTTATGTGTT 2049
Db 268 ValValIleAspAsnArgIleSerCysHisArgIleLeuIlePheGluGlyValAlaPheIle 287
2050 GTAGGATAGACAAACAGGGGATCCTGTCAACGAGGGCTTAAATTTGAAGGCGCTTTAAA 2109
QY 288 TyrIleuMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
2110 TATATAAATGGGTCAATATGAAATTTGACATCAAGTGTGGAAGGCTCAATATCTTACCTTCT 2169
QY 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyIleTyr 327
2170 CGATATGATTTTCATTGACTTGAATGTGTGGGCAATCCAGCGTGTCTCTATGAGAGATAC 2229
QY 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
2230 CTCTCCCTGATGAGCTATATGACAGGTACAGATATCTTCAAGGTTGCTATTGCTGGGCCC 2289

QY 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
Db 2290 CCACTGACTCTGTGATCTCTATGATATACAGATACAG 2328
RESULT 3
US-10-170-789-39
Sequence 39, Application US/10170789
Publication No. US20030180930A1
GENERAL INFORMATION:
APPLICANT: Rachel E. Meyers
APPLICANT: Olandt, Peter J.
APPLICANT: Kapeller-Libermann, Rosana
APPLICANT: Curtis, Rory A. J.
APPLICANT: Williamson, Mark
APPLICANT: Welch, Nadine
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
FILE REFERENCE: 10448-191001
CURRENT APPLICATION NUMBER: US/10/170,789
CURRENT FILING DATE: 2002-06-13
PRIOR APPLICATION NUMBER: US 09/797,039
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/06525
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/186,061
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 09/882,166
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: PCT/US01/19269
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/212,078
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 09/934,406
PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: PCT/US01/26052
PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: US 60/226,740
PRIOR FILING DATE: 2000-08-21
PRIOR APPLICATION NUMBER: US 09/861,801
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: PCT/US01/16549
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: US 60/205,508
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: US 09/801,267
PRIOR FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: PCT/US01/07138
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 60/187,454
PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/829,671
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: PCT/US01/40483
PRIOR FILING DATE: 2001-04-11
PRIOR APPLICATION NUMBER: US 60/197,508
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 09/961,721
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: PCT/US01/29904
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 60/235,023
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US 10/045,367
PRIOR FILING DATE: 2001-11-07
PRIOR APPLICATION NUMBER: US 60/246,561
PRIOR FILING DATE: 2000-11-07
PRIOR APPLICATION NUMBER: US 09/801,275
PRIOR FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: PCT/US01/07074
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 60/187,420
PRIOR FILING DATE: 2000-03-07

```
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-170-789-39

Alignment Scores:
Pred. No.:      9,81e-226      Length:      2649
Score:          1836.50       Matches:      360
Percent Similarity: 67.54%    Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match:     95.01%      Indels:      173
DB:              12          Gaps:         1

US-10-070-464-7 (1-360) X US-10-170-789-39 (1-2649)

QY      1 GIUGIUAAPALARGSERIAGLYVALATHPHEVAILEUGINGIUGIUPHEASPAR 20
DB      730 GAAGAAGATCCAGATCAGCTGAGAGCTGCTACCTTTGTTCTCCAGAAAGATTGATGA 789
QY      21 TYRSEGIYTYRTTPTCPYSPROLYSALAGIUTHRTHRPROSERGIYGIYLSILEU 40
DB      730 TATCTGCTATTGGTGGTGTCCAAAGCTGAACCACTCCAGTGGTGAATAATCTT 849
QY      41 ARGILELEUTYRGIUGIUNASPGIUSERGIUVALGIUILLIHSIVALTHRSEPRO 60
DB      850 AAAAAATCTATATGAAGAAAATGATGATCTGAGGTGGAATATATCATCTACATCCCT 909
QY      61 METLEUGIUTHRARGIRGALASPSEPHARGTYRPROLYSTHRGIYTHRALAASPRO 80
DB      910 ATCTTGAAACAAGAGGAGGAGATTCATTCCTGTTATCCAAAACAGGTACAGAAATCT 969
QY      81 IYSAVALTHRPHLYSMETSERGIUILEMETILEASPAIAGIUGIYARGILEILE 98
DB      970 AAAAGTCACTTTTAAGATGTCAGAAATAATGATGATGCTGAAGAGATCATAGATGTC 1029
QY      98 ----- 98
DB      1030 ATATGATAGGAATTAATTAACCTTTTGAGATTCTATTGAAAGAGTTGAATATATGTC 1089
QY      98 ----- 98
DB      1090 AAGAGTGATGAGATCTCTGAGGAAATATGCTTGTCATCTACTAGATGCTCCAG 1149
QY      98 ----- 98
DB      1150 ACTCGCTGACAGATAGTGTGATCTGACCTGAATATTATCCACAGTAGAAGATGATT 1209
QY      98 ----- 98
DB      1210 ATGGAAAGCAGAGACTCATTTGAGTCAIGTCCGATTTCTGTAGCGCACTAATTATCTAT 1269
QY      98 ----- 98
DB      1270 GAAGAAACAACAGACATCTGATATATATCATGATCTTTTCATGTTTTCCTCCAAAGT 1329
QY      98 ----- 98
DB      1330 CACGAAGAGAAATGAGTTTATTTTGCTCTGAATGCAAAAACAGGTTTCGTCATTTA 1389
QY      98 ----- 98
DB      1390 TACAAAATTACATCTATTTTAAAGAAAGCAATATTAAGATCCAGTGTGGCTGCCCT 1449
QY      98 ----- 98
DB      1450 GCTCAAGTGAATTCAGTGTCTCTATCAAGAAGAGATGCAATTACAGTGTGATGGA 1509
QY      99 ----- ValASPGIUVAlARGLIUEUValIYR 107
DB      1510 GAAGTTCTTGCGCGCATGATCTAATATCCAAAGTGAAGTCAAGAGCTGGATATAT 1569
```

```
QY      108 PHEGIUGIYTHRILEASPSERPROLEUGIUNHSILEUVALIASETYRVALASN 127
DB      1570 TTGAAGGACACCAAGACTCCCTTTAAGACATCCTGTAGTGTGATGATACGTAAAT 1629
QY      128 PROGIUGIUVAlTHRARGLEUTHRASPARGIYTYRSETHISESCYSCYSLIESEGIN 147
DB      1630 CCTGAGAGGTGACAGAGGTGATGATCCGTGGCTACTACATTTCTTCTGATCATGTCAG 1689
QY      148 HISYASPEPHEPHEILESERIYTYRSEANGLIYASNPROIHCYVALSERLEU 167
DB      1630 CACTGTGACTTCTTTATAGTATGATTAACAGAGAAATCCACACTGTGTGCTCCCTT 1749
QY      168 TYRYSLEUSERSEPROGIUNASPSPROCTHRCYSLYETHLYGIUNPHEPVALTHR 187
DB      1750 TACAAAGCTATCAAGTCCGAAAGATACCCACTTGCAAAAACAAGGAATTTGGGCACCC 1809
QY      188 ILEUASPSERIALGIYPROLEUPROASPIYTRHPRPROGIUUILEPHESERPHEGIU 207
DB      1810 ATTTGGATTGACGAGGTCTCTCTCTGACTATACCTCCAGAAATTTCTCTTTGAA 1869
QY      208 SEPTHRTYRGIYTHRILEUTHRYGLIYMETLEUTYRYSPROHISASPLEUGINPROGIY 227
DB      1870 AGTACTAGTGAATTTACATTTGATGAGTGTCTCTCAACCTCATGATCTACAGCCGTGA 1929
QY      228 IYSLYTYRPROTHRVALLEUPHEILETYRGIYGIYPROGIYVALGINLEUVALASASN 247
DB      1930 AAAAAATCTCTACTGTGTGCTGTCTATATATGATGTGTCTCCAGAGTGCAGTTGGTGAATAT 1989
QY      246 ARGPHLYSGIYVALIYTYRPHEARGLUEANTRHLEUALASERLEUGIYTYRVAL 267
DB      1990 CGGTTTAAAGAGTCAAGATTTCCGCTTGAAATACCTTAGCTCTTAGATTATGATGTT 2049
QY      268 VALIYVAlIIEASPSANARGIYSERCYSHISARGIYLEULYSPHEGIUGIYALAPHELYS 287
DB      2050 GTAGGTATGACAAACAGGGGATCTGTCCACCAAGGCTTAAATTTGAAGCGGCTTTAAA 2109
QY      288 TYRYSMETGIYGINIIEGIUUILEASPSAGIUNVALGIUGIYLEUGIYTRHLEUALASER 307
DB      2110 TATAAATAGGTGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2169
QY      308 ARGTYRASPHEILEASPLEUSPARGYVALIYIIEHSIGIYTRPSETYRGIYGIYTYR 327
DB      2170 CGATATGATTTTATTGATGATTAATCGTGTGGGATCCAGCGGTGCTTATGAGGATAC 2229
QY      328 LEUSERLEUMETALALEUMETGLIARGSERASPILLEPHARGYVALIIEALAGIYALA 347
DB      2230 CTCTCCCTGATGCAATTAATGACAGGTGAGATATCTTCAAGGTTCTATTTGCTGGGCC 2289
QY      348 PROVALTHRLEUTRIPLEPHETYRASPTRHGIYTYRTHR 360
DB      2290 CAGTCACTCTGTGATCTTCTATGATACAGATACAGC 2328

RESULT 4
; Sequence 2, Application US/0976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Junien, Jean-Louis
; APPLICANT: Riviere, Pierre
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2671
; TYPE: DNA
```

ORGANISM: Homo sapiens
US-09-976-674-2

Alignment Scores:

Pred. No.:	9,946-226	Length:	2671
Score:	1836.50	Matches:	360
Percent Similarity:	67.54%	Conservative:	0
Best Local Similarity:	67.54%	Mismatches:	0
Query Match:	95.01%	Indels:	173
DB:	10	Gaps:	1

US-10-070-464-7 (1-360) x US-09-976-674-2 (1-2671)

```
OY      1  GUGUASPAALAArSerAlaGlyValAlaThrPheValLeuGlnGlnuPheSpArg 20
Db      737  GAAGAAGATGCCAGATCGAGCTGAGTCGCTACCTTGTTCTCCAGAAAGATTGATAGA 796
OY      21  TyrSerGlyTyrTrpCysProGlyAlaGluTrpThrProSerGlyGlyValLeu 40
Db      797  TATTCTGCTATTGGTGGTGTCCAAAGCTGAACACTCCAGTGGTGTAATAATCTT 856
OY      41  ArgIleuTyrGlnGluuAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
Db      857  AGAATTCTATATGAGAAATATGATCTGAGTGAATATTATTCATGTTACATCCCT 916
OY      61  MetLeuGluTrpArgAlaAspSerPheArgIleProLysThrGlyThrAlaAsnPro 80
Db      917  ATGTGGAAACAGAGAGGCGACATTCATTCCTTATCTCTAAACAGGTACACCAATCCT 976
OY      81  LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
Db      977  AAAAGTCACTTTTAAAGATGTCAGAAATTAATGATTGATGCTGAAGAGAAATCATAGATGTC 1036
OY      98  ----- 98
Db      1037  ATAGATAAGAACTAATTCAACTTTTGAGATTCTATTGAAGAGTTGAATATATATGCC 1096
OY      98  ----- 98
Db      1097  AGAGCTGATGAGCTCTGAGGAGAAATATGCTTGTCATCTACTAGATGCTCCAG 1156
OY      98  ----- 98
Db      1157  ACTGCCTGCAGATAGTGTGATCTCACTGAATTAATTATCCACAGTAGAAGATGATT 1216
OY      98  ----- 98
Db      1217  ATGGAAGGAGAGACTCATGATGATGAGTCTGATCTGTGACGCCACTAATATATCTAT 1276
OY      98  ----- 98
Db      1277  GAAGAAACAGACAGATCTGATTAATATCCATGACATCTTTCATGTTTTTCCCAAGT 1336
OY      98  ----- 98
Db      1337  CACGAAGAGAAATTGAGTTATTTTGGCTCTGATGCAAAACAGGTTCCGTCAATTTA 1396
OY      98  ----- 98
Db      1397  TACAAATATACATCTATTTTAAAGAAAGCAATATAAACATCCAGTGGTGCGCT 1456
OY      98  ----- 98
Db      1457  GCTCAAGTATTCATGATGTCCTATCAAGAAGAGATAGCAATTACAGTGTGATAGG 1516
OY      99  -----ValAspGluValArgArgLeuValTyr 107
Db      1517  GAAGTCTTGGCCGCGATGATCTATATATCCAAAGTTGATGAAGTCAAGAGCTGGTATAT 1576
OY      108  PheGlnGlyThrLysAspSerProLeuGlnHisIleuTyrValValSerTyrValAsn 127
Db      1577  TTTGAAGGACCAAGAGACTCCCTTTAGAGCATCACTGTAGTGTAGTACGTACGTAAAT 1636
```

```
OY      128  ProGlnGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysAlaSerGln 147
Db      1637  COTGAGAGGGTACAGAGGCTGACGACCGTGCTACTACATCTTTCGTCAGTCAG 1696
OY      148  HisCysAspPhePheIleSerIleTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db      1697  CACTGTACTCTTTATATAGTAGATATGTAACCGAAGAAATCCACACTGTGTCCCTT 1756
OY      168  TyrLysLeuSerSerProGluuAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
Db      1757  TACAAAGTATCAAGTCTTGAGATGACCCCAACTTCCAAACAAAGAAATTTTGGGCACC 1816
OY      188  IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGlnuIlePheSerPheGlu 207
Db      1817  ATTTGGATTACAGCAGGCTCTCTTCTTCACTACTCCCGAAATTTTCTTTTGA 1876
OY      208  SerThrTrpGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuInProGly 227
Db      1877  AGTACTACTGATTTTACATTTGATATGAGATGCTCTACAGCTCATGATTTACAGCTGGA 1936
OY      228  LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
Db      1937  AAGAAATATCTCACTGCTGCTGTTATATATGTTGCTCAGGTGCAAGTTGATGATAT 1996
OY      268  ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db      1997  CGATTTAAAGAGTCAAGTATATTTCCGCTTGAATACCTTAAGCTCTCTAGTTATGAGT 2056
OY      268  ValValIleAspAsnArgGlySerCysHisArgGlyLysLeuLysPheGlnGlyAlaPheLys 287
Db      2057  GTAGTGATAGACAAAGGAGATCTGTACACGAGGCTTTAAATTTGAAGGCCCTTTAA 2116
OY      288  TyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTyrLeuAlaSer 307
Db      2117  TATAAAAGGGTCAATATGAATATGACGATCAGTGAAGAAAGCTCCAAATATCTAGCTTCT 2176
OY      308  ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyTyr 327
Db      2177  CGATATGATTTCAATGACTATGATGCTGTGGCATCCAGCGTGGCTTCTATGAGCATAC 2236
OY      328  LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyVal 347
Db      2237  CTCTCCCTGATGCAATATATGAGAGGTGATATCTTCAAGGATGCTATATCTGGGGCC 2296
OY      348  ProValThrLeuThrIlePheTyrAspThrGlyTyrThr 360
Db      2297  CCAATCACTCTGTGATCTTCTATGATACAGATACAG 2335

RESULT 5
US-10-170-789-37
; Sequence 37, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Welch, Nadine
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
; FILE REFERENCE: 10448-191001
; CURRENT APPLICATION NUMBER: US/10/170,789
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/797,039
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06525
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186,061
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 09/882,166
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19269
```

```

/ PRIOR FILING DATE: 2001-06-15
/ PRIOR APPLICATION NUMBER: US 60/212,078
/ PRIOR FILING DATE: 2000-06-15
/ PRIOR APPLICATION NUMBER: US 09/934,406
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: PCT/US01/26052
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: US 60/226,740
/ PRIOR FILING DATE: 2000-08-21
/ PRIOR APPLICATION NUMBER: US 09/861,801
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: PCT/US01/16549
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: US 60/205,508
/ PRIOR FILING DATE: 2000-05-19
/ PRIOR APPLICATION NUMBER: US 09/801,267
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07138
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,454
/ PRIOR FILING DATE: 2000-03-07
/ PRIOR APPLICATION NUMBER: US 09/829,671
/ PRIOR FILING DATE: 2001-04-10
/ PRIOR APPLICATION NUMBER: PCT/US01/40483
/ PRIOR FILING DATE: 2001-04-11
/ PRIOR APPLICATION NUMBER: US 60/197,508
/ PRIOR FILING DATE: 2000-04-18
/ PRIOR APPLICATION NUMBER: US 09/961,721
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: PCT/US01/29904
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: US 60/235,023
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US 10/045,367
/ PRIOR FILING DATE: 2001-11-07
/ PRIOR APPLICATION NUMBER: US 60/246,561
/ PRIOR FILING DATE: 2000-11-07
/ PRIOR APPLICATION NUMBER: US 09/801,275
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07074
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,420
/ PRIOR FILING DATE: 2000-03-07
/ NUMBER OF SEQ ID NOS: 63
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 37
/ LENGTH: 3143
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (229)...(2874)
US-10-170-789-37

```

Alignment Scores:

```

Pred. No.: 1,29e-225 Length: 3143
Score: 1836.50 Matches: 360
Percent Similarity: 67.54% Conservative: 0
Best Local Similarity: 67.54% Mismatches: 0
Query Match: 95.01% Indels: 173
DB: 12 Gaps: 1

```

US-10-070-464-7 (1-360) x US-10-170-789-37 (1-3143)

```

QY 1 GIUGLHASPAAIARGSERIAGLYVALAATHPhValleuGInguInguInpheaSPaYg 20
DB 958 GAAGAAGATGCCGATCACTGAGTCGCTAACCTTGTCTCCAAAGAATTGATAGA 1017
QY 21 TYrSeGlyTYrTYrTYrTYrTYrTYrTYrTYrTYrTYrTYrTYrTYrTYrTYrTYr 40
DB 1018 TATTCGGCTATTCGGCTATTCGGCTATTCGGCTATTCGGCTATTCGGCTATTC 1077
QY 41 ArgIleuTYrGInguInguInguInguInguInguInguInguInguInguInguIn 60

```

```

DB 1078 AGAATTCATATGAAAGAAATGATATCTGAGGGGAAATATTCATGTTACATCCCT 1137
QY 61 MetIeugIuThrArgArgAlaSPSerPheaTYrProLYsThrGlyThraIAsnPro 80
DB 1138 ATGTTGAAACAAAGAGGAGGAGATTCATTCCTGTTCTCTAAACAGGTACAGCAATCCT 1197
QY 81 LysValIThrPheIysMetSerGInuIleMetIleAspAlaGluTYrArgIleIle 98
DB 1198 AAAGTCACCTTTTAAATGATGCAAAATAATGATGATGCTGAAGAGATCATATGATGTC 1257
QY 98 ----- 98
DB 1258 ATGATTAAGAACTAATTCACCTTTTGAGATTCATTTGAAGAGATTGATATATGTC 1317
QY 98 ----- 98
DB 1318 AGAGCTGATGGAATCTCTGAGGAAATATGCTGTGTCATCTACTAGATGCTCCAG 1377
QY 98 ----- 98
DB 1378 ACTGCTGACAGATGATGTTGATCTACCTGAATTAATTTATCCAGTGAAGATGATGTT 1437
QY 98 ----- 98
DB 1438 ATGAAAGCAGAGACTCATTTGATGACGTCCTGATTTCTGACCCCAATATATCTAT 1497
QY 98 ----- 98
DB 1498 GAAGAAACAAACAGACATCTGATTAATATCCATGACATCTTCATGTTTCCCAAGT 1557
QY 98 ----- 98
DB 1558 CACGAAAGAAATGAGTTATTTTGGCTCTGAATGCAAAACAGTTCCGTCAATTA 1617
QY 98 ----- 98
DB 1618 TACAAATTAATCATCTATTTTAAAGAAACAAATTAACGATCCAGTGGGCTGCT 1677
QY 98 ----- 98
DB 1678 GCTCCAGATGATTTCAAGTGTCTATCAAGAGAGATGCAATTACAGTGTGAATG 1737
QY 99 ----- 99
DB 1738 GAAGTTCTTGCCCGGATGATCTATATCAAGTTGATGAGTCAAGAGCTGATAT 1797
QY 108 PheGluGlyThrLYsSPSerProLeuGInuIshIseuTYrValIseTYrValAsn 1857
DB 1798 TTTGAAGCACCAGAGCTCCCTTTAGAGCATCCTGTAGCTAGTACGTTAAT 1857
QY 128 ProGlyGluValIThrArgLeuThraSPaTYrTYrTYrTYrTYrTYrTYrTYrTYr 1917
DB 1858 CCGGAGAGGTGACAGAGCTGACCTGACCTGCTACTCACTTCTTGACATCACTCAG 1917
QY 148 HisCysASPhePheIleSerLYsTYrSerAsnGInLYsAsnProHisCysValSerLeu 1977
DB 1918 CACTGTGATCTTTTATAGTATGATATTAACCAAGAAATCAACACTGTGTCTCT 1977
QY 168 TYrLYsLeuSerSerProGInuSPaSPProThrCysLYsThrLYsGInuPheTYr 1977
DB 1978 TACAAGCTATCAAGTCTGAAAGATGACCAACTGCGAAACAAAGAAATTTGGCCACC 2037
QY 188 ILeuAspSerIleGlyProLeuProAspTYrThrProProGInuIlePheSerPheGIn 207
DB 2038 ATTTTGATTCAGCAGAGTCTCTCTGACTATACCTCCAGAAATTTCTTTTGAA 2097
QY 208 SerThrThrGlyPheThrLeuTYrGlyMetLeuTYrLYsProHisAspLeuGInProGly 227
DB 2098 AGTACACTGAGATTTATGATGAGGAGCTCTAACAACCTCATATGATCAGCCTGGA 2157
QY 228 LysLYsTYrProThrValleuPheIleTYrGlyGlyProGInuValGInleuValAsn 247

```

```

Db      2158 AAGAAATATCCACTGCTGCTTTCATATATGCTGCTCAGGTGCAAGTGGTAATAT 2217
Qy      248 ArgPheIysGlyValIysTyrPheArgLeuAnthrLeuAlaSerLeuGlyTyrValVal 267
Db      2218 CGGTTAAAGAGGATCAAGATATTCGCTGAAATACCTAGCTCTAGTATATGCTT 2277
Qy      268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db      2278 GTAGGATAGACACAGGGGATCCGTGACCGAGGGCTTAAATTTGAAGCGCCCTTTAA 2337
Qy      288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrPheAlaSer 307
Db      2338 TATATAATGGGTCAATATGAAATTTGACGATCAGGTGGAAGGACTCCAAATATCTAGCTT 2397
Qy      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 327
Db      2398 CGATATGATTTTCATGACTTAGATCGTGTGGGCATCCAGGCTGGTCTTATGAGAGATAC 2457
Qy      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db      2458 CTCTCCCTGATGGCATTTATGACAGAGTCAGATATCTTCAGGGTGTGCTATTCTGGGGCC 2517
Qy      348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
Db      2518 CCAATCACTCTGTGATCTTCTATGATACAGATACAG 2556

RESULT 6
US-09-976-674-12
; Sequence 12, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Jumeau, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 4829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-12

Alignment Scores:
Pred. No.: 9,6e-223 Length: 4829
Score: 1816,50 Matches: 360
Percent Similarity: 67,29% Conservative: 0
Best Local Similarity: 67,29% Mismatches: 0
Query Match: 93,97% Indels: 175
Db: 10 Gaps: 1

US-10-070-464-7 (1-360) x US-09-976-674-12 (1-4829)
Qy      1 GluGluAspAlaArgSerIleGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
Db      943 GAAAGAGATGCCAGATCAGCTGAGTCCGCTACTTTTCTCCAGAGAAATTTGATAGA 1002
Qy      21 TyrSerGlyTyrTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
Db      1003 TATTCCTGCTATTTGCTGTGCTCCAAAGCTGAACACACTCCAGTGTGTAATAATTTCT 1062
Qy      41 ArgIleLeuTyrGluGluAspAspGluSerGlyValGluIleIleHisValThrSerPro 60
Db      1063 AGAATTCATATGAAGAAATATGATGATCTGAGGTGAATATTCATGATCCCT 1122
Qy      61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db      1123 ATGTTGAAACAGAGAGGAGATTCATTCCTTATCCAAACAGGTACAGCAATTCCT 1182
Qy      81 LysValThrPheLysMetSerGluIleMetIleAspAlaGlyArgIleIle----- 98
Db      1183 AAAGTCACCTTTAAAGATGTCAGAAATATATGATTTGCTGAAGAGAGATCATATATGTC 1242
Qy      98 ----- 98
Db      1243 ATATGATAGAGACTAATTCACCTTTTGATTCATTTTGAAGAGGTGAATATATTTGCC 1302
Qy      98 ----- 98
Db      1303 AAGCTGATGATCCTCTGAGGAAATATGCTGTGTCATCTACTAATGCTCCAG 1362
Qy      98 ----- 98
Db      1363 ACTGCGCTACAGATATGTTGTATCTCACCTGAATTATTTATCCAGTATAGAATATGTT 1422
Qy      98 ----- 98
Db      1423 ATGAAAGCAGAGACTATGATGATCAGTCTGATTCGTGACGCCACTAATATATAT 1482
Qy      98 ----- 98
Db      1483 GAAAGAAACACAGACATCGATTAATATATCATGACATCTTCATGTTTCCCAAGT 1542
Qy      98 ----- 98
Db      1543 CACGAAAGAGAAATTGAGTTTATTTTGGCTCTGAATGCAAAACAGGTTCCGTCATTTA 1602
Qy      98 ----- 98
Db      1603 TACAAATATACATCTATTTTAAAGAAACAAATATTAACGATCCAGTGTGGCTGCT 1662
Qy      98 ----- 98
Db      1663 GCTCCAGATGATTCAAGTGTCTCTTCAAAAGAGATACGAATTAACAGTGTGAATGG 1722
Qy      99 -----ValAspGluValArgArgLeuValTyr 107
Db      1723 GAAGTTCTTGGCCGGCATGATCTAATATCAAGATTTGATGAAGTCAAGAGGCTGATAT 1782
Qy      108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
Db      1783 TTTGAAGGACCAAAAGACTCCCTTTAGGACATCACTGTACTGATCTGATTAATCTTAAT 1842
Qy      128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
Db      1843 CCTGAGAGGTGACAAAGCTGTGACTGACCGGTGCTACTCACATTTCTTGCTGATCAGTCA 1902
Qy      148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db      1903 CACTGTGACTTCTTTATATAGTATATGATTAACGAAACATCCACACTGTGTCTCTT 1962
Qy      168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
Db      1963 TACAAGCTATCAAGCTCGAAGATGACCCCACTTGCAAAACAAAGGAATTTTGGGCCAC 2022
Qy      188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207
Db      2023 ATTTTGGATTCAGCAGGTCTCTTCTGATCTACTCTCTCCGAAATTTTCTTTTGA 2082
Qy      208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db      2083 AGTACTACTGATGATTTACATATGATGAGATGCTCTCAAGCCCTCATATCTACAGCTGGA 2142
Qy      228 LysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnLeuValAsn 247
Db      2143 AAGAAATATCTCACTGTGCTGTTCATATATGATGTGTCTCTCAAGGCGAGTGTGTAATA 2202
Qy      247 snArgPheLysGlyValIysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValI 267
Db      2203 ATCGGTTAAAGAGATCAAGTATTTCCGCTGAATACCTTAGCCTCTTAGGTATATGTGG 2262
```


QY 267 alValValIleaspasnarArglySerCysHisArgGlyLeuLysPheGluGlyAlaPheL 287
 Db 2263 TTGTAGTATAGAACACAGGGGATCCTGTCACCGAGGGCTTAATTGAGCGGCTTTA 2322
 QY 287 ySTyLyMeGlyGlnIleGlnIleaspasGlnValGlnGlyLeuGlnTyPheLysAlaS 307
 Db 2323 AATATAAATGGGTCAAAATAGAAATAGACATCAGGTGAGAGACTCCAAATATCTAGCTT 2382
 QY 307 eTArgTyAspPheIleaspLeuaspArgValGlyIleHisGlyTyPheSerTyArgIlyT 327
 Db 2383 CTCGATATGATTTTCATTGATCTTAGATCGTGTGGGATCCAGCGGTGCTCTATGAGGAT 2442
 QY 327 yTLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 347
 Db 2443 ACCCTCCCTGATGGCATTAATGACAGAGTCAGAAATCTTCAGGGGTGCTATTGCTGGGG 2502
 QY 347 laPheValThiLeuTrpIlePheTyAspTrpGlyTyThr 360
 Db 2503 CCCAGTCACTCTGTGATCTTCTATGATACAGGATACAG 2543

RESULT 7

US-09-976-674-22
 ; Sequence 22, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akinsanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Junien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: US 60/240,117
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 22
 ; LENGTH: 4685
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-976-674-22

Alignment Scores:

Pred. No.: 8,42e-185 Length: 4685
 Score: 1522.00 Matches: 312
 Percent Similarity: 58.54% Conservative: 0
 Best Local Similarity: 58.54% Mismatches: 1
 Query Match: 78.74% Indels: 221
 DB: 10 Gaps: 2

US-10-070-464-7 (1-360) x US-09-976-674-22 (1-4685)

QY 1 GIuGIuaspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheaspArg 20
 Db 943 GAAGAAGATGCCAGATCACTGAGTCCGCTACCTTGTCTCCAGAAATAATTGATAGA 1002
 QY 21 TySerGlyTyTrpTyPheProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
 Db 1003 TATTCTGGCTATTGGGTGTGCCAAAGCTGAACAACACTCCAGGTGGTGAATAATCTT 1062
 QY 41 ArgIleLeuTyGlnGluLysaspGluSerGluValGlnIleIleHisValThrSerPro 60
 Db 1063 AGAATTCATATAGAAATAATGATGATCTGAGGTGGAATATTCATATCATTCCTCT 1122
 QY 61 MetLeuGluThrArgArgAlaaspSerPheArgTyProLysThrGlyThrAlaaspPro 80
 Db 1123 ATGTTGGAACAACAGAGGCGCATTCATCCGTTATCCATAAAGGTACACCAATCTT 1182
 QY 81 LysValThrPheLysMetSerGluIleMetIleaspAlaGluGlyArgIleIle----- 98
 Db 1183 AAGTCACTTTTAAGATGTCAGAAATTAATGATTGATGTGAAGAGATCATAGATGTC 1242

QY 98 ----- 98
 Db 1243 ATAGATAAGAACTAATTCAACTTTGAGATTCTATTGAAAGATTGAATATATTGCC 1302
 QY 98 ----- 98
 Db 1303 AGAGCTGATGACCTCTGAGGGAATAATGCTTGCTGATCTTACTAGATGCTCCAG 1362
 QY 98 ----- 98
 Db 1363 ACTGCTACAGATAGTGTGATCTCACCTGAATATTATTATCCAGTAGAAGATGATGTT 1422
 QY 98 ----- 98
 Db 1423 ATGAAAGACAGACATCATAGTCAAGTCCGTGATCTGTGAGCCCACTAATATCTAT 1482
 QY 98 ----- 98
 Db 1483 GAAAGAACACAGACATCTGATTAATATCCATGACATCTTTCATGTTTCCCAAGT 1542
 QY 98 ----- 98
 Db 1543 CACGAAGAAATAGATTATTTTGGCTGTGATGCAAAACAGTTCCGTGATTTA 1602
 QY 98 ----- 98
 Db 1608 TACAAATTAATCATCTATTATTAAGAAACAAATATTAACGATCCAGTGGTGGCTGCT 1662
 QY 98 ----- 98
 Db 1663 GCTCAAGTATTCAGTGTCTTCAAGAGGATAGCAATTAACAGTGTGATAG 1722
 QY 99 -----ValaspGlnValArgLeuValIly 107
 Db 1723 GAAGTCTTGGCCGCATGATCTAATATCCAAAGTGAATGAAAGTGGTGGATAT 1782
 QY PheGluGlyThrLysaspSerProLeuGlnHisIleuTyValValSerTyValaen 127
 Db 1783 TTTGAAGGACCAAGAGACTCCCTTTAAGACATCCGTAGCGTACGTACGTACGTAAAT 1842
 QY 128 ProGlyGlnValThrArgLeuThrAspArgGlyTySerHisSerCysCysIleSerGln 147
 Db 1843 CCTGAGAGGTGACAGAGCTGACGACCGTGGCTACTGACATCTTCTGCGATCAGTCAG 1902
 QY 148 HisCysaspPhePheIleSerLysTySerAsnGlnLysAsnProHisCysValSerLeu 167
 Db 1903 CACTGTGACTCTTTATATAGTAAGTATGTAACCAAGAAATCCACACTGTGTGCTCTT 1962
 QY 168 TyLysLeuSerSerProGlnaspAspProThrCysLysThrLysGluPheTrpAlaThr 187
 Db 1963 TACAAAGCTATCAAGTCTGAAAGATGACCACTTGCACAAACAAAGAAATTTGGGCCACC 2022
 QY 188 IleLeuaspSerAlaGlyProLeuProaspTyThrProProGluIlePheSerPheGlu 207
 Db 2023 ATTTCGATTGAGT----- 2036
 QY 208 SerThrThrGlyPheThrLeuTyGlyMetLeuTyLysProHisaspLeuGlnProGly 227
 Db 2036 ----- 2036
 QY 228 LysLysTyProThrValLeuPheIleTyGlyGlyProGlnValGlnLeuValAsnAsn 247
 Db 2037 -----CTCAGTCAAGTGGTGAATAAT 2060
 QY 248 ArgPheLysGlyValLysTyPheArgLeuAsnThrLeuAlaSerLeuGlyTyValVal 267
 Db 2061 CGGTTTAAGGAGCAAGATATCCGCTGTAATACCTAGCCTCTCTAGTTATGTGTT 2120
 QY 268 ValValIleaspasnarArglySerCysHisArgGlyLeuLysPheGlnGlyAlaPheLys 287
 Db 2121 GTAGTATAGACACAGGGGATCTGTACCGAGGGCTTAATTTGAAGGCGCTTTAAA 2180

```
QY      288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
      |||
Db      2181 TATATAATGGGTCAAAATAGAAATTGACGATCAGGTGGAAAGACTCCAAATTCTAGCTTCT 2240

QY      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyValTyr 327
      |||
Db      2241 CGATATGATTCTTATGCTAGATCGTGGGCAATCCAGCGCTGCTCTATGGAAGGATAC 2300

QY      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
      |||
Db      2301 CTCTCCGATGGCATTTAATGACAGAGTCAGATATCTTCAGGGTTGCTATGCTGGGGCC 2360

QY      348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
      |||
Db      2361 CCAGTCACCTGTGGATCTTCTATGATACAGATACACG 2399

RESULT 8
US-09-976-674-20
; Sequence 20, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junier, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 4676
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-20

Alignment Scores:
Pred. No.:      9.74e-185      Length:      4676
Score:          1521.50      Matches:      309
Percent Similarity: 57.97%      Conservative: 0
Best Local Similarity: 57.97%      Mismatches: 1
Query Match:      78.71%      Indels:      224
DB:              10          Gaps:      2

US-10-070-464-7 (1-360) x US-09-976-674-20 (1-4676)

QY      1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValIleuGlnGluIleuPheAspArg 20
      |||
Db      943 GAAGAAGATGCGCAGATCAGCTGAGTGCCTACCTTGTCTCCAGAGAAATTGATAGA 1002

QY      21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrProSerGlyGlyLysIleu 40
      |||
Db      1003 TATTGCGCTATTGGGTGTCCAAAGCTGAAACCACTCCAGTGTGGTAAATTTCT 1062

QY      41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
      |||
Db      1063 AGAATCTATATAGAAAGAAATGATGATCATCGAGTGGAAATTAATCATGTCATCCCT 1122

QY      61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
      |||
Db      1123 ATGTGGAAACAAAGAGGCGCAGATTCATTCCTTAATCCCTAAACAGGTACGCAAAATCCT 1182

QY      81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyValGlyIleIle 98
      |||
Db      1183 AAAGTCACTTTTAAAGATGTCAGAAATATATGATGAGCTGAAAGAGATCATGATGTC 1242

QY      98 ----- 98
Db      1243 ATGATAAGAACTAATTCACCTTTTGAGATTCATTTGAAGAGTTGAATATATGCGC 1302
```

```
QY      98 ----- 98
Db      1303 AGAGCTGATGACTCTCTGAGGAAATATGCTTGGTCCATCTACTAGATCGCTCCAG 1362
      |||
QY      98 ----- 98
Db      1363 ACTGCGCTACAGATAGTGTGATCTGACCTGAATTTATATCCAGTAGAAGATGATGTT 1422

QY      98 ----- 98
Db      1423 ATGAAAGGAGAGACTCATGATGAGTCCCTGATTCGTGAAGCCACTAATATCTAT 1482

QY      98 ----- 98
Db      1483 GAAGAAACAAACAGACATCTGATATAATATCATGACATCTTCATGTTTTCCTCCAAAGT 1542

QY      98 ----- 98
Db      1543 CACGAAGAGAAATGAGTTTATTTTGGCTCTGAATGCMAAACAGGTTCCGTCAATTTA 1602

QY      98 ----- 98
Db      1603 TACAAATTAATCATCTATTTTAAAGAAAGCAATATTAACATCCAGTGGTGGCTGCT 1662

QY      98 ----- 98
Db      1663 GCTCCAAAGTATTTCAAGTGTCTCTATCAAGAGAGATAGCAATTACAGTGTGAATGG 1722

QY      99 ----- ValAspGluValArgArgLeuValTyr 107
      |||
Db      1723 GAAGTCTTGGCCGGCATGATCTAAATATTCGAATGATAGTAAAGCAAGAGCTGGTATAT 1782

QY      108 PheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValIleSerTyrValAsn 127
      |||
Db      1783 TTGAAGGACCAACAAAGACTCCCTTTAGAGCATCACTGTACGTAGCATTTACGTAAT 1842

QY      128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSerGln 147
      |||
Db      1843 CTGGAAGAGGTGACAAAGCTGACCTGACCGGTGCTACTCATCTTCTGCTGATCAGTCAG 1902

QY      148 HisCysAspPhePheIleSerLysTyrSerAspGlnLysAsnProHisCysValSerLeu 167
      |||
Db      1903 CACTGAGACTCTTTTAAAGTAAAGTATAGTAAACAGAAAGATCCACCTGTGTCTCT 1962

QY      168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
      |||
Db      1963 TACAGCTATCAAGTCTGAAAGATGACCCAACTGCAAAACAAAGAAATTTGGCCAC 2022

QY      188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207
      |||
Db      2023 ATTTGGATTCACACAGTCTCTCTTCCGACTATATCTCTCCAGAAATTTCTTTGAA 2082

QY      208 SerThrThrGlyPheThrLeuTyrArgIleMetLeuTyrLysProHisAspLeuGlnProGly 227
      |||
Db      2083 AGTACTACAGATTATCATATGATGATGATCTCTACAGGCTCTATGATCTACAGCTGGA 2142

QY      228 LysLysTyrProThrValIleuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
      |||
Db      2143 AAGAAATATCTTACTGTGCGTTCATATAGTGGTGTG----- 2180

QY      248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
      |||
Db      2180 ----- 2180

QY      268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
      |||
Db      2180 ----- 2180

QY      288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
      |||
Db      2181 -----GTCAAATAGAAATTTGACGATCAGGTGGAAAGACTCCAAATTTCTAGCTTCT 2221

QY      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyValTyr 327
```

```

Db      2232 CGATTGATTTCATTGACTAGATCGTGGGATCCAGCGCTGGCTTATGAGGATAC 2291
Qy      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db      2292 CTCCTCCCTGATGACATTAAAGCAGAGGTGAGATATCTTCAGGGGTTCATTGCTGGGGCC 2351
Qy      348 ProValThrLeuThrIlePheThrAspThrGlyTyrThr 360
Db      2352 CCAGTCACCTCTGTGGATCTTCTATGATACAGGATACAG 2390

RESULT 9
US-09-976-674-8
; Sequence 8, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 4523
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-8

Alignment Scores:
Pred. No.: 4,69e-147 Length: 4523
Score: 1229.00 Matches: 258
Percent Similarity: 48.41% Conservative: 0
Best Local Similarity: 48.41% Mismatches: 1
Query Match: 63.58% Indels: 275
DB: 10 Gaps: 2

US-10-070-464-7 (1-360) x US-09-976-674-8 (1-4523)
Qy      1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
Db      943 GAAGAGATGCCAGATCAGCTGAGTGGTACCTTCTCTCCAGAGAAATTTGATAGA 1002
Qy      21 TyrSerGlyTyrTyrTyrCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
Db      1003 TATTTCTGGCTATGTGGTGTCTCCAAAGCTGAAACAACTCCAGTGTGTAAATTTCTT 1062
Qy      41 ArgIleLeuTyrGluGluAspGluSerGluValGluIleIleHisValIleThrSerPro 60
Db      1063 AGAATTCATATAGAGAAATAGATGATCTGAGGTGAAATATATCTCATGTATACATCCCT 1122
Qy      61 MetLeuGlnThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAspPro 80
Db      1123 ATGTGGAACCAAGAGAGGAGAGATTCATTCGTTATCTTAAACAGGTACAGCAATCCT 1182
Qy      81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
Db      1183 AAAGTCACTTTTAAGATGTCAGAAATATATGATGATCTGAGAGAGATCATATAGATGTC 1242
Qy      98 ----- 98
Db      1243 ATGATRAGGAACTAATTCAACCTTTGAGATTCTATTGAGAGCTTGAAATATATTGCC 1302
Qy      98 ----- 98
Db      1303 AGAGCTGATGACTCCTGAGGAAATATGCTTGTCATCTACTAGATCGCTCCAG 1362
Qy      98 ----- 98

```

```

Db      1363 ACTCGCTACAGATAGTGTGATCTCACCGTAATTATTCACAGTAGAATATGTT 1422
Qy      98 ----- 98
Db      1423 ATGAAAGCAGAGACTCATGATGACGTGCTGATTTCTGTGACGCCACTAATTATCTAT 1482
Qy      98 ----- 98
Db      1483 GAAAGAAACACAGACATCTGATTAATATTCATGACATCTTTCATGTTTTTCCCAAGT 1542
Qy      98 ----- 98
Db      1543 CACGAGAGAAATGATGTTTATTTTGGCTGTGAATGCAAAACAGGTTTCCGTCATTTA 1602
Qy      98 ----- 98
Db      1608 TACAAATTATCATCTATTTTAAAGAAAGCAATATTAACGATCCAGTGTGGGCTGCC 1662
Qy      98 ----- 98
Db      1663 GCTCCAGATGATTTCAAGTGTCTCTATCAAAAGAGAGATAGCAATTACAGTGTGAATGG 1722
Qy      99 ----- ValAspGluValArgArgLeuValTyr 107
Db      1723 GAAGTCTTGCCCGCATGATCTATATATCCAAAGTTGATGAGAGCTGAGATAT 1782
Qy      108 PheGluGlyThrLysAspSerProLeuGlnHisIleLeuTyrValValSerTyrValAsn 127
Db      1783 TTTGAAGGACCAAAACCTCCCTTTAGAGCAACACCTGTACGTAGTCATTCGTAAAT 1842
Qy      128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
Db      1843 CCTGAGAGGTGACAAAGCTGACTGACCGGTGACTACATCTTCTGTCATCAGTCAG 1902
Qy      148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db      1903 CACTGAGACTTCTTTATAGTAGTAGTAGTAAACCAAGAAATCCACACTGTGTGCCCTT 1962
Qy      168 TyrIleLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPalaThr 187
Db      1963 TACAACTATCAAGTCTGTGAAGTAGACCAACTGTGCAAAACAAAGAAATTTGGGCCAAC 2022
Qy      188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db      2023 ATTTGATTCACAGAGTCTCTCTGATATATCTCCAGAAATTTCTCTTTGAA 2082
Qy      208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db      2083 AGTACTACTGATTTTACATGATGAGGATGCTCTACAGGCTCATATCTTACAGCCCGGA 2142
Qy      228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
Db      2143 AAGAAATATCTTCTGTGCTGTTCAATATATGAGGT----- 2178
Qy      248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db      2178 ----- 2178
Qy      268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db      2178 ----- 2178
Qy      288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db      2178 ----- 2178
Qy      308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyGlyTyr 327
Db      2178 ----- 2178
Qy      328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347

```

Db 2179 -----CGGTT-GCTATTGCTGGGCC 2198
Qy 348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
Db 2199 CCAAGTCACTGTGTGATCTTTGATGATACAGGATACAG 2237

RESULT 10
US-09-976-674-4
; Sequence 4, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2617
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-4

Alignment Scores:
Pred. No.: 2,466-146 Length: 2617
Score: 1220.50 Matches: 236
Percent Similarity: 54.31% Conservative: 54
Best Local Similarity: 44.19% Mismatches: 69
Query Match: 63.14% Indels: 175
Gaps: 2

US-10-070-464-7 (1-360) x US-09-976-674-4 (1-2617)

Qy 2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluGluPheAspArgTyr 21
Db 710 GATGACCCCAAGTGTGGGGGTGGGCCACCTTCGTATACAGGAAGATTGCACCGCTTC 769
Qy 22 SerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
Db 770 ACTGGGTACTGTGTGGTCCCAACAGCTCTCGGAAGGTTCAGAGGGCTCAAGAGCGCTG 829
Qy 41 ArgIleLeuTyrGluGluAspGluSerGluValGluIleIleHisValThrSerPro 60
Db 830 CGAATCCCTGTATGAGGAAGTCAATGATGCCAGGTGGAGGTCAATTCAGTCCCTCTCT 889
Qy 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db 890 GCGCTAGAAAGAAAGAAAGACGAGCTCGATCGGTACCCAGAGACAGGACGAAGAATCCC 949
Qy 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
Db 950 AAGATTGCTTGAACCTGCTGAGTTCCAGACTACACCCAGGCGGCAAGATGCTTCGACC 1009
Qy 98 ----- 98
Db 1010 CAGAGAAAGAGAGCTGTGTGACGCCCTTCAGCTCCGTTCCCGAAGGTGGATGACATGCC 1069
Qy 98 ----- 98
Db 1070 AGGGCCGGGTGGACCCCGGATGCAAAATACGCTGGGCCATGTTCTGSAACCGGCCAG 1129
Qy 98 ----- 98
Db 1130 CAGTGGCTCAGCTGTCTCTCTCCCGCGCCCTGTTCATCCGAGAGACAGAGANTGAG 1189
Qy 98 ----- 98

Db 1190 GAGCAGCGGCTAGCCTCTGCCAGAGCTGTCCCAAGAAATGTCACGCCGTATGTGTTAC 1249
Qy 98 ----- 98
Db 1250 GAGAGGTCAACCAAGTGTGATCATGTTCATGACATCTTCTATCCCTTCCCAATCA 1309
Qy 98 ----- 98
Db 1310 GAGGAGAGACAGAGCTGTCTTCTTCCGCGCATATGATGACAGACCGGCTTTCGCAT 1369
Qy 98 ----- 98
Db 1370 TTGTACAAAGTCACCGCGTTTAAATCCAGGGCTACGATTGAGTGAGCCCTTACAC 1429
Qy 98 ----- 98
Db 1430 CCGGGGAAGATGAATTAAAGTCCCACTTAAGAGAGATGCTCTGACCGCGGTGA 1489
Qy 99 -----ValAspGluValArgLeuVal 106
Db 1490 TGGAGGTTTGTGCGAGGACAGGCTCAAGATCTGGTCAATGAGAGACCAAGCTGTGTG 1549
Qy 107 TyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValIleSerTyrVal 126
Db 1550 TACTTCAGGGCAACAGACACCGCTGAGACACCTCTTACGTGTGCTGCTATGAG 1609
Qy 127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSer 146
Db 1610 GCGCGCGCGAGATGTGATCGCTCACACAGCCCGCTCTCCATAGCTGCTCCATGAGC 1669
Qy 147 GluHisCysAspPhePheIleSerLysTyrSerAsnGluLysAsnProHisCysValSer 166
Db 1670 CAGAACTTCGACATGTTGTGTGACGACATACAGACGTTGAGACACCGCGCTGGGAC 1729
Qy 167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluThrTrpAla 186
Db 1730 GCTCAAGCTGAGGGGCGCCGACGACGACCCCTTGACACAGACACCCCGCTTGTGGGT 1789
Qy 187 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe 206
Db 1790 AGCATGATGAGAGCGACCGACGCTGCGCCCGGATTTATGTTCTCCAGAGATCTTCATTTC 1849
Qy 207 GluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnPro 226
Db 1850 CACAGCGCTCGAATGTGCGGCTCTACGGCATGATCTACAGGCCCAAGCCCTTGACGCA 1909
Qy 227 GlyLysLysTyrProThrValLeuPheIleTyrGlyLysProGluValGluLeuValAsn 246
Db 1910 GGGAAAGAGCACCCACGCTCTTGTATATGAGAGGCCCCCAAGGTGCAGCTGTGAT 1969
Qy 247 AsnArgPheLysGlyValIleTyrThrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266
Db 1970 AACTCCTTCAAAGCATTAATATCTGCGGCTCAACACACTGTGCTCTCGGCTACGCGC 2029
Qy 267 ValValValIleAspAsnArgLysCysHisArgGlyLeuLysPheGluValAlaPhe 286
Db 2030 GTGGTTGTGATTGACGAGGAGGCTCTCTGACGAGGAGGCTTCGGTGTGAAGGGCCCTG 2089
Qy 287 LysTyrLysMetGlyGluIleGluIleAspAspGluValGluGlyLeuGlyThrLeuAla 306
Db 2090 AAAAACCAATGGGCGAGGTGAGATTCAGGACCAAGGTGAGGGGCTCAGATTGCTGGCC 2149
Qy 307 SerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGly 326
Db 2150 GAGAGATATGCTTCATCATGACCTAGCGCAATTCGATTCATTCATGCTGCTCTACGCGG 2209
Qy 327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346
Db 2210 TTCTCTCTGCTCATGAGGGCTAATCAACAGCCCGAGGTTCAGAGTGGCATGCGGGGT 2269
Qy 347 AlaProValThrLeuThrIlePheThrTyrAspThrGlyTyrThr 360
Db 2270 GCCCGGTACCGTGTGATGGCTTACGACACAGGGTACCT 2311


```

; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patent version 3.1
; SEQ ID NO: 24
; LENGTH: 4302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-24

Alignment Scores:
Pred. No.: 5,4e-146 Length: 4302
Score: 1220.50 Matches: 236
Percent Similarity: 54.31% Conservative: 54
Best Local Similarity: 44.19% Mismatches: 69
Query Match: 63.14% Indels: 175
DB: 10 Gaps: 2

US-10-070-464-7 (1-360) x US-09-976-674-24 (1-4302)

QY 2 GIUAPPALAARXSERIAGLYVALAATHRPHLEVALLEUGLNUHLEUPHESAPRGTYR 21
DB 1066 GATGACCCCAAGCTGCGGGGTGGCCACCTTGATATACAGAAAGTTCCAGCGCTTC 1125
QY 22 SERGIYTYRTRIPCYSPROLYSALAGIUNTRTHPRROSERGIYGLY--LYSILEU 40
DB 1126 ACTGGTACTGGTGGTGGCCCAAGCTCTCGGAAAGTTTCAGAGSGCTCAAGACGCTG 1185
QY 41 ARGILEUITYRGILUNASAPGUSERGIUVALGLUILELLEHISVALTHRSEPRO 60
DB 1186 CCAATCTGTATAGAGAAATCATGATGATCGAGGTGAGGTCAATTCAGCTCCCTCTCCT 1245
QY 61 METLEUGLUNTHARARGLALASPSERPHARGTYRPROLYSTHGLYTHRALASPRO 80
DB 1246 GCGCTAGAAAGAAAGAGACGACTGTATCGGTACCCCAAGACAGCAGCAAGAAATCCC 1305
QY 81 LYSVALTHRPHLEYSERGIUULEMETILEASPALAGIUGLYARGILELLE----- 98
DB 1306 AAGATTGCTTGAAATGCGCTAGTTCCAGACTGACAGCCAGGCAAGATCGCTCGACC 1365
QY 98 ----- 98
DB 1366 CAGGAGAAAGAGCTGTGACAGCCCTTCACTCGCTGTCCCAAGAGTGAGATACGCGC 1425
QY 98 ----- 98
DB 1426 AAGGCGCGGTGACCCCGGATGGCAAAATACGCTGGGCAATGTTCTTGACCGGCCCGC 1485
QY 98 ----- 98
DB 1486 CAGTGGCTCAGCTGTCTCTCTCCCGCGCCGCTGTGATCCCGAGGCAAGAGATGAG 1545
QY 98 ----- 98
DB 1546 GAGCAGCGCTAGCCTCTGCAAGACTGTCCCGAGAAATGTCCAGCCGATGTGTGTAC 1605
QY 98 ----- 98
DB 1606 GAGGAGGTACCAACGCTTGATCAATGTTCAATGACATCTTATCCCTTCCCAATCA 1665
QY 98 ----- 98
DB 1666 GAGGAGAGAGAGAGCTGTCTTCTCCGCGCAATGATGAAGCAAGCCGCTTCTGCAAT 1725

```

```

QY 98 ----- 98
DB 1726 TTGTCAAAGTACCGCGCTTTTAAATCCAGGCTACGATGTAGTGAGCCCTTACG 1785
QY 98 ----- 98
DB 1786 CCGGGGGAATGAAATTAAAGTCCCAATTAAAGAAAGATTGCTTGACAGCGGTGAA 1845
QY 99 -----ValASPGLUVALARGARGLEUVAL 106
DB 1846 TGGAGGTTTGGCCAGGACCGCTCCAAAGATCTGGGTCAATGAGAGACCAAGCTGGTG 1905
QY 107 TYRPHLEGLIYTHRLYSASPSERPROLEUGLHSHISLEUTRYVALISERTYRVAL 126
DB 1906 TACTTCCAGGGACCAAGGACACGCGCTGGAGACACCACTTCACTGTGTCACTATGAG 1965
QY 127 ASPNROGLYGLUVALTHRARGLEUHRASPARAGIYTYRSETHISERCYSLIESER 146
DB 1966 GCGGCGCGGAGATGATACGCTTCAACGACCGCCGCTTCCCATAGCTGTCCATGAGC 2025
QY 147 GINH4CYASPHEPHEILESERIYTSERASNGILYASNPROIHISCYSVALSER 166
DB 2026 CAGAACTTGCATATGTTCTGTACGACCACTACAGAGGTAGACAGCGCCCTCGTGAC 2085
QY 167 LEUTYRLYSEUSERPROGLUNASAPPROTHRYSLEYTHRLYSGLUHPHETPALA 186
DB 2086 GTCTCAAGCTAGCGCGCCGCCACGACGACCCCTTGCAAGAGAGCCCGCTTCTGGGCT 2145
QY 187 THRILEUASPSERLAGLYPROLEUPROAPRYRTHPRPROGLUILEPHESERPE 206
DB 2146 AGCATGATGAGAGCGCCAGCGTCCCGCGGATATGTTCTTCCATAGATCTTCAATTC 2205
QY 207 GUSERTHRTHGLYPHERTHREUTRYRGIMETLEUTRYLYSPROHISASPLEUGINPRO 226
DB 2206 CACAGCGCGTCCGAGTGTGGGCTTACGCGCAATGATACAGGCCCAAGCCTTGACAGCA 2265
QY 227 GLYLYSLYSTRPROTHRYVALLEUPHEILETYRGILYGLYPROGLINVALGILNLEUVALASN 246
DB 2266 GGGAAAGAGCACCCACCGCTCTTGTATATGAGGCGCCCGAGGTGACGCTGTGCAAT 2325
QY 247 ASPARGPHEIYGLYVALLYSTRPHEARGLEUASPTHLEUALASERLEUGIYTRYVAL 266
DB 2326 AACTCTTAAAGGATCAAGTACTGCGGCTTCAACACTGGCTCCCTCGGCTACGCC 2385
QY 267 VALVALVALILEASPARAGIYSERCYSHISATRGILEUYSAPHEGLIUGLYALAPHE 286
DB 2386 GGGGTGTGATGACGCGGAGGCTCTGTCAAGCAGGAGGCTTCGTTCAAGGCGCCCTG 2445
QY 287 LYSYRLYSMETGLYGLNILEGLUILEASPARGLINVALGLIYLEUGINTYRLEUALA 306
DB 2446 AAAAACCAATATGGCCAGGTGAGATCGAGACAGGTGAGGCGCTTGAGTGTGTGGCC 2505
QY 307 SERATYRASPHEILESPLEUSPARGLYLYLEHISGLYTRPSETRYRGILY 326
DB 2506 GAGAGTATGCTTCACTCACTGACCGAGTTCATGACATGAGCTGTGTCAGGGAGGCG 2565
QY 327 TYRLEUSERLEUMETALAEUMETGINRSEASPDILEPHEARGVALALALEAGIY 346
DB 2566 TTCTCTGCTCATAGGCGCTAATCCAAAGCCCGAGGTTCCTCAAGGTGCGCATCGGGGT 2625
QY 347 ALAPROVALTHLEUTRPILEPHERYRASPTRYRTHR 360
DB 2626 GCCCGGTACCGCTGTGATGCTTACGACACAGGAGGTACACT 2667

RESULT 13
US-09-976-674-36
; Sequence 36, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis

```

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patentin version 3.1
SEQ ID NO 36
LENGTH: 1180
TYPE: DNA
ORGANISM: Homo sapiens
US-09-976-674-36

Alignment Scores:
Pred. No.: 1.69e-143 Length: 4180
Score: 1201.00 Matches: 234
Percent Similarity: 55.09% Conservative: 53
Best Local Similarity: 44.91% Mismatches: 72
Query Match: 62.13% Indels: 162
DB: 10 Gaps: 3

US-10-070-464-7 (1-360) x US-09-976-674-36 (1-4180)

QY 2 GLUAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArgTyr 21
DB 1066 GATGACCCCAAGCTGCGGGTGGCCACCTTCTGATACAGAAAGTTCCAGCCGCTTC 1125
QY 22 SerGlyTyrTrpTyrCysProIysAlaGluThrProSerGlyGly--LysIleLeu 40
DB 1126 ACTGGGTACTGGTGGTGGCCCAAGCCTCTCGGAAAGTTCAAGAGGCGCTCAAGAGCGCTG 1185
QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
DB 1186 CGATCTCTGATGAGAGAGTCGATGAGTCGAGGTGAGGTCACTTCCGCTCTCTCT 1245
QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProIysThrGlyThrAlaAsnPro 80
DB 1246 GCGCTGAAAGAAAGAAAGAGAGAGCTGATCGGTACCCAGAGAGCGAGCAAGAAATCCC 1305
QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
DB 1306 AAGATTGCTTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1365
QY 101 GluValArgArgLeu----- 105
DB 1366 CAGGAGAGAGAGCTGTGTGACAGCCCTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1425
QY 105 ----- 105
DB 1426 AGGCGCGGTGACCCGGGATGGCAATACGCTGGGCCATGTTCTGTGACCGGCCCCAG 1485
QY 105 ----- 105
DB 1486 CAGTGGCTCCAGCTGCT 1545
QY 106 -----ValTyr 107
DB 1546 GAGCAGCGCTAGCT 1605
QY 108 PheGlu----- 109
DB 1606 GAGAGAGTCAACAAGCTGTGATCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1665
QY 109 ----- 109
DB 1666 GAGGAGAGAGAGAGCTGTGCTTTCTCCGCGCAATGAATGAAGCAAGCCGCTTTCAT 1725
QY 109 ----- 109
DB 1726 TTGTACAAAGTCAAGCCGCTTTAAATCCAGAGGCTAGCATTTGAGTAGAGCCCTTCAGC 1785
QY 109 ----- 109

DB 1786 CCGGAGAAAGATGAATTTAAGTCCCATTTAAGAAAGACATTCTGTACACCGGTGAA 1845
QY 110 -----GlyThrLysAspSerProLeuGlnHis 119
DB 1846 TGGAGAGTTTGGCGAGCAGCGGCTCCAGAGGACCAAGAGACAGCGCTGAGACAC 1905
QY 120 LeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgIleTyr 139
DB 1906 CTCTAGCTGTGATGATGAGCGCGCGCGAGATGTGATGATGATGATGATGATGATGATGAT 1965
QY 140 SerHisSerCysCysIleSerGlnHisCysAspPheIleSerTyrSerAsnGln 159
DB 1966 TCCCATAGCTGTCTCCATGAGCAAGCTTCGATGATGATGATGATGATGATGATGATGAT 2025
QY 160 LysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCys 179
DB 2026 AGACAGCCCGCTGCTGACAGCTCAAGCTCAAGCTGAGCGGCCCAAGACAGACCCCTGAC 2085
QY 180 LysThrLysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThr 199
DB 2086 AAGCAGCCCGCTTCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2145
QY 200 ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 219
DB 2146 CTTCAGAGATCTTCATTTCCACAGCGCTGAGATGAGCTTCAAGCATGATCTAC 2205
QY 220 LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGly 239
DB 2206 AAGCCCAAGCTTGTGACAG 2265
QY 240 ProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThr 259
DB 2266 CCCCAGGTGAGTGTGATTAATCTTCAAGAGATCAAGTGTGCGCTCAACACA 2325
QY 260 LeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCysHisArgGly 279
DB 2326 CTGGCTCTCTGGGCTGACCGCTGTTGATTTGAGCGAGGAGGAGCTCTCTCTCTCTCTCT 2385
QY 280 LeuLysPheGluGlyValAlaPheLysTyrLysMetGlyGlnIleGlnIleAspAsnGlnVal 299
DB 2386 CTTCGTTGAGAGGAGGCTGAGAAACCAATGAGGAGAGTGAATGAGAGAGAGAGAGAG 2445
QY 300 GlnGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
DB 2446 GAGGCGCTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2505
QY 320 HisGlyTyrSerTyrGlyTyrLysSerLeuMetAlaLeuMetGlnArgSerAspIle 339
DB 2506 CATGCTGTCTCTACGGGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2565
QY 340 PheArgValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyr 359
DB 2566 TTCAAGGTGGCCATCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2625
QY 360 Thr 360
DB 2626 ACT 2628

RESULT 14
US-09-976-674-34
Sequence 34, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akimsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117

```

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 4263
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-34

Alignment Scores:
Pred. No.:      1,74e-143      Length:      4263
Score:          1201.00        Matches:      334
Percent Similarity: 55.09%      Conservative: 53
Best Local Similarity: 44.91%      Mismatches: 72
Query Match:      62.13%      Indels:      162
DB:               10          Gaps:          3

US-10-070-464-7 (1-360) x US-09-976-674-34 (1-4263)

QY      2  GluSepAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArgTyr 21
      1066 GATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCAATACAGAGAGAGTTCGACCGCTTC
QY      22 SerGlyTyrTrpTyrProGlyValAlaGluThrThrProSerGlyGly---LysIleLeu 40
      1126 ACTGGGTACTGGTGTGCTGCCACAGCTCTCGGAGAGTTCCAGAGGCTCTCAAGAGCTG 1185
QY      41 ArgIleLeuTyrGluGluAlaAspGluSerGluValGluIleLeuIleValThrSerPro 60
      1186 CGAATCCCTGTATGAGAGAGTGCATAGTCCAGAGTGGAGGTCAATTCAGTCCCTCTCT
QY      61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAspPro 80
      1246 GCGCTAGAGAAAGAGAGACGAGACTCGTATCGGTACCCAGAGACAGCAGCAAGATCCC 1305
QY      81 LysValThrPheLysMetSerGlyIleMetIleAspAlaGluGlyArgIleLeuValasp 100
      1306 AAGATTGCTTGAATCTGGTGAAGTTCAGACTGACACGAGGAGCAAGATCGTTCGACC 1365
QY      101 GluValArgArgLeu-----ValTyr 107
      1366 CAGAGAGAGAGAGCTGTGTCAAGCCTTCAGCTCGCTGTTCCGAAAGTGAATACATCCGC 1425
QY      105 -----ValTyr 107
      1426 AGGCGCGGGTGAAGCCCGGAGATGCAATACGCTGGGCCATGTCTGGACCGGCCAG 1485
QY      105 -----ValTyr 107
      1486 CAGTGGCTCCAGCTGTCTCTCTCCCGCGCTGTTCCATCCGAGACAGAGAAATGAG 1545
QY      106 -----ValTyr 107
      1546 GAGCAGCGGCTAGCCTCTGCGCAGAGCTGTCCCGCAGAAATGTCCAGCCGTATGTGTAC 1605
QY      108 PheGlu-----ValTyr 107
      1606 GAGGAGTTCACCAAGTGTGATCAATGTATGACATCTTATCCTTCCCCCAATCA 1665
QY      109 -----ValTyr 107
      1666 GAGGAGAGAGAGAGCTGTCTTTCGCGCCCAATGAATGACAGACCGGCTTCTGCAT 1725
QY      109 -----ValTyr 107
      1726 TTGTACAAAGTCAACCGCGTTTAAATCCAGGGCTACGATTTGAGTGAAGAGCCCTTACGC 1785
QY      109 -----ValTyr 107
      1786 CCGGAGAGAGATTAATTAAGTCCCATTAAGAGAGATTTGCTGTACCAAGCGGTGA 1845
QY      110 -----ValTyr 107
      110 -----ValTyr 107

```

```

DB      1846 TGGAGGTTTGGCGAGGACAGGCTCCAGGGCACCAAGACACCGCCGCTGAGACACAC 1905
QY      120 LeuTyrValIleSerTyrValAlaAspProGlyGluValThrArgLeuThrAspArgIleTyr 139
      1906 CTCTACGTGTGACGTATGAGCGCGCCGAGATGTATGCTTACACACCGCCGCTTC 1965
QY      140 SerHisSerCysValIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGln 159
      1966 TCCCATAGCTGCTCCATAGGACAGAACTTCGACATGTCTGTACGACATACAGACGCGT 2025
QY      160 LysAsnProHisCysValSerLeuTyrTyrLysLeuSerSerProGluAspAspProHisCys 179
      2026 AGCAGCGCGCCCTGTGTCACGTCTACAGCTGAGCGGCCCGCAGACACACCCCTGCAC 2085
QY      180 LysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThr 199
      2086 AAGCAGCCCGCTTGTGGCTAGCATGATGAGAGAGCAGCAGTCCCGCCCGATTTATGTT 2145
QY      200 ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 219
      2146 CTTCCAGAGATCTTCATTTCCACAGCGCTCGATGTGCGGCTCTACGCGCATGATCTAC 2205
QY      220 LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly 239
      2206 AAGCCCAAGCCTTGTACAGCCAGAGAGAGACCCACCGCTCTTGTATATGAGAGC 2265
QY      240 ProGlnValGlnLeuValAlaAspArgPheLysGlyValLysTyrPheArgLeuAsnThr 259
      2266 CCCCAGGTGACAGCTGTGAATTAATCTCTTCAAGAGCATCAAGTACTTGGGCTCAACACA 2325
QY      260 LeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGly 279
      2326 CTGGCTCTCCCTGAGGCTACCGGTGTGTATGACGAGAGGCTCTGTCAGAGAGG 2385
QY      280 LeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnVal 299
      2386 CTTCGTTTGAAGGGGCTCTGAAAAACCAATGGCCAGGTGTGAGATGAGAGACACAGG 2445
QY      300 GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
      2446 GAGGCTCTGACGTTGTGTGCGCAGAGATGTGCTTCAATGACCTGAGCCGAGTTGCCATC 2505
QY      320 HisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle 339
      2506 CATGGCTGTCTTACGGGGGCTTCTCTGCTCAAGGGGCTATCCACAGCCCGCAGGTG 2565
QY      340 PheArgValAlaIleAlaGlyAlaProValThrLeuTyrPheTyrAspThrGlyTyr 359
      2566 TTCAAGTGTGCATGTGCGGGTGTCCCGGTACCGTGTGATGTGCTTACGACAGAGGTAC 2625
QY      360 Thr 360
      2626 ACT 2628
DB      2626 ACT 2628

RESULT 15
US-10-098-841-83
; Sequence 83, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Jiao-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie

```


/ APPLICANT: Qian, Xiaohong B.
 / APPLICANT: Dmanac, Radoje T.
 / TITLE OF INVENTION: No. US20020197679A1 Nucleic Acids and
 / FILE OF INVENTION: Polypeptides
 / FILE REFERENCE: 784CIP2
 / CURRENT APPLICATION NUMBER: US/10/098,841
 / PRIOR FILING DATE: 2002-03-13
 / PRIOR APPLICATION NUMBER: 09/598,042
 / PRIOR FILING DATE: 2000-06-20
 / PRIOR APPLICATION NUMBER: 09/552,317
 / PRIOR FILING DATE: 2000-04-25
 / PRIOR APPLICATION NUMBER: 09/488,725
 / PRIOR FILING DATE: 2000-01-21
 / NUMBER OF SEQ ID NOS: 331
 / SOFTWARE: pc_files Version 1.0
 / SEQ ID NO: 83
 / LENGTH: 3262
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: (121)..(2334)
 / US-10-098-841-83

Alignment Scores:

Pred. No.:	2 05e-141	Length:	3262
Score:	1183.50	Matches:	230
Percent Similarity:	55.34%	Conservative:	55
Best Local Similarity:	44.66%	Mismatches:	65
Query Match:	61.23%	Indels:	165
DB:	13	Gaps:	3

US-10-070-464-7 (1-360) x US-10-098-841-83 (1-3262)

QY 2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluPheAspArgTyr 21
 DB 499 GATGACCCCAAGCTGCGGAGTGTGCGCACCTTCGTATACAGGAAGATTGCGACCGCTTC 558
 QY 22 SerGlyTyrTrpProPysProLysAlaGluThrThrProSerGlyLys---LysIleLeu 40
 DB 559 ACTGGGTACTGGTGGGCGCCACAGCTCTCTGGGAAGTTCAAGAGGCGCTCAAGACGCTG 618
 QY 41 ArgIleLeuTyrGlnGluAsnAspGlnSerGluValGlnIleHisValThrSerPro 60
 DB 619 CGATCTGTATAGAGAAAGTCATGATCGAGGTGAGGTCAATTACGTCCTCTCTCT 678
 QY 61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
 DB 679 GCGCTAGAAAGAAAGGAGACGACTGCTATCGGTACCCAGGACAGGACAGCAAGATATCCC 738
 QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGlnGlyArgIleIle----- 98
 DB 739 AAGATTGCTTGAATCTGGCTGAGTTCACAGCTGACAGCCAGGCGCAAGATGCTCTGACC 798
 QY 98 ----- 98
 DB 799 CAGGAGAAGAGCTGTGACAGCCCTTCACTGCTGTCCCGAAGGTGAGTATCATCGCC 858
 QY 98 ----- 98
 DB 859 AAGGCGGAGTGAACCGGAGTGGCAAAATACGCTGGGCAATGTTCTTGAGACGGGCCCGAG 918
 QY 98 ----- 98
 DB 919 CAGTGGCTCCAGCTGCT 978
 QY 98 ----- 98
 DB 979 GAGCAGCGGCTAGCT 1038
 QY 98 ----- 98
 DB 1039 GAGGAGGTCAACCACTGTGATCAATGTTTATGATGATCTTATCTCTCTCTCTCTCTCTCT 1098

QY 98 ----- 98
 DB 1099 GAGGAGAGAGACAGCTCTGCTTTCTCCGCCAATGATGACAGACCGGCTTCTGCAAT 1158
 QY 98 ----- 98
 DB 1159 TTGTACAAAGTACCGCGCTTTAAATATCCAGGGCTACGATGAGTACGACCTTCACG 1218
 QY 99 ----- ValAspGluValArgArgLeu 105
 DB 1219 CCGGGGAAGGTGAGAGAGAGCTTACGATGCTATGTGATGATAGAGAGACCAAGCTG 1278
 QY 106 ValTyrPheGlnGlyThrLysAspSerProLeuGlnHisIleLeuTyrValValSerTyr 125
 DB 1279 GTGTACTTCCAGGAGACCAAGACACCGCGCTGGAGACACCTCTACGTTGTCAGCTAT 1338
 QY 126 ValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIle 145
 DB 1339 GAGCGCGCGCGAGATCGTACGCTTACACAGCGCGGCTTCTCCATAGCTGCTCCATG 1398
 QY 146 SerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysVal 165
 DB 1399 AGCCGAACTTGCAGCATGTTGTCAGCCACTACAGCAGCGTGAAGACCGCGCGCTGCGT 1458
 QY 166 SerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyr 185
 DB 1459 CAGCTTACACCTGACGCGCGCCGACGACGACCGCTTCACAGAGAGCGCGCTCTGCG 1518
 QY 186 AlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGlnIlePheSer 205
 DB 1519 GCTAGCATATGAGAGACGCT-----AAGATCTTCCAT 1551
 QY 206 PheGlnSerThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGln 225
 DB 1552 TTCCACACCGCGCTGATGTCGCGCTCTACGCGCATGATCTACAGGCCCAAGCGCTTGCAG 1611
 QY 226 ProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuVal 245
 DB 1612 CCAAGGAAGAACACCCCAAGCTCTTGTATATGAGAGGCCCCAGGATGACGCTGCTG 1671
 QY 246 AsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyr 265
 DB 1672 AATTACTCTTCAAAAGCATCAAGTACTTGGCGCTCAACACTGGCTCTCTCTGCGCTAC 1731
 QY 266 ValValValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGlnGlyAla 285
 DB 1732 GCGGTGTGTGATTGACCGGACGAGGCTCTCTGTCAGCGAGGCTTCTGATTCAGAGGGGCC 1791
 QY 286 PheLysTyrLysMetGlyGlnIleGlnIleAspAsnGlnValGlnGlyLeuGlnTyrLeu 305
 DB 1792 CTGAAACCAATATGGCCAGGATGAGATCGAGACAGGTGAGGAGGCTGCAATTCGATG 1851
 QY 306 AlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGly 325
 DB 1852 GCCGAGAAGTATGGCTTATGACCTGACCTGAGCCGAGTTCATCATGCTGCTGCTGCTG 1911
 QY 326 GlyTyrLysSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAla 345
 DB 1912 GGCCTTCCTCTCTCATGAGGAGTAAATCCACAGCCCAAGTGTCAAGGTGACCATCGCG 1971
 QY 346 GlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThr 360
 DB 1972 GGTGCCCGGTACCGCTGTGATGAGCTTACGACAGAGGTTACCT 2016

Search completed: October 16, 2003, 09:10:05
 Job time : 309.152 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:24:12 ; Search time 2162.86 Seconds

(without alignments)
4045.398 Million cell updates/sec

SUMMARIES

0 29 gb_gse2.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFLQEEFDR.....RVAAIGAPVLTWTFYDTGYT 360
Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+pn.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10070464/runat.15102003.113553.24829/4dp.query.fasta.1.2652
-DB=EST -OPMT=fastdp -SUFFIX=est -MINMATCH=0.1 -IOCPCT=0 -IOCPXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=Pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NOM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10070464@cgn2.1.8056@runat.15102003.113553.24829 -NCPU=6 -ICPU=3
-NO MMAP -LARGEBUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -NARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hiv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mus:*
23: em_gss_mam:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1.*

Result No.	Score	* Query Match	Length	DB	ID	Description
1	1756.5	90.9	3143	11	AK016546	AK016546 Mus muscu
2	1669	86.3	5517	11	AK029788	AK029788 Mus muscu
3	1415	73.2	1042	12	BM557438	BM557438 AGENCOURT
4	1325.5	68.6	910	13	BO675606	BO675606 AGENCOURT
5	1297	67.1	957	13	BO675606	BO675606 AGENCOURT
6	1237.5	64.0	1041	12	BI084090	BI084090 AGENCOURT
7	1232	63.7	746	14	BY751026	BY751026 BY751026
8	1205	62.3	968	13	BO671635	BO671635 AGENCOURT
9	1204.5	62.3	3327	11	AK050021	AK050021 Mus muscu
10	1204.5	62.3	3327	11	AK050021	AK050021 Mus muscu
11	1204.5	62.3	3457	11	AK078301	AK078301 Mus muscu
12	1122	58.0	1082	13	BU239476	BU239476 60332338
13	1117	57.8	674	10	BE888665	BE888665 601513061
14	1108.5	57.3	1090	12	BM910838	BM910838 AGENCOURT
15	1099	56.9	639	10	BI655205	BI655205 602344074
16	1089	56.3	753	9	AL040398	AL040398 DKFZ4334A
17	1057	54.7	658	10	BE388695	BE388695 601286345
18	991	51.3	831	12	BE974324	BE974324 602844075
19	976.5	50.5	853	10	BF694553	BF694553 602081767
20	970.5	50.2	898	13	BU0908783	BU0908783 AGENCOURT
21	951.5	49.2	603	13	BU483718	BU483718 603470021
22	949	49.1	789	10	BE709118	BE709118 602675382
23	937	48.5	982	13	BO678015	BO678015 AGENCOURT
24	922	47.7	897	13	BU191638	BU191638 AGENCOURT
25	919	47.5	695	10	BE715866	BE715866 602675282
26	916.5	47.4	853	12	BI223892	BI223892 602941035
27	905	46.8	820	13	BU522948	BU522948 AGENCOURT
28	902.5	46.7	845	13	BU216915	BU216915 602092373
29	884.5	45.9	896	12	BE921463	BE921463 602824365
30	868	44.9	788	10	BI616397	BI616397 602338360
31	866	44.8	871	13	BU196183	BU196183 AGENCOURT
32	863	44.6	821	12	BI690774	BI690774 603143392
33	833	43.1	931	13	BU235022	BU235022 603409569
34	826	42.7	479	12	BI135177	BI135177 UI-M-BH3
35	788.5	40.8	1021	12	BE974587	BE974587 602844820
36	787	40.7	687	14	CB851806	CB851806 UI-CF-FW0
37	765	39.6	713	14	CA321029	CA321029 UI-M-FW0
38	762	39.4	432	13	BY520490	BY520490 BY520490
39	762	39.4	957	13	BU234147	BU234147 603412033
40	756	39.1	741	14	CB595102	CB595102 AGENCOURT
41	753	39.0	551	9	AL643656	AL643656 AL643656
42	746	38.6	735	13	BU447277	BU447277 603766685
43	738	38.2	428	13	BY548432	BY548432 BY548432
44	736	38.1	1023	13	BO681703	BO681703 AGENCOURT
45	735	38.0	1080	10	BE203783	BE203783 601868764

ALIGNMENTS

RESULT 1
LOCUS AK016546
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length cDNA library, clone:4932434f09 product:DIPEPTIDYL PEPTIDASE 8 homolog
[Homo sapiens], full insert sequence.
ACCESSION AK016546
VERSION AK016546.1 GI:12855334
KEYWORDS HTG, CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

JOURNAL MEDLINE
99279253
PUBMED 10349636
REFERENCE

2
AUTHORS

3
TITLE

4
JOURNAL
MEDLINE
PUBMED
REFERENCE

5
AUTHORS

6
TITLE

7
JOURNAL
MEDLINE
PUBMED
REFERENCE

8
AUTHORS

9
TITLE

10
JOURNAL
MEDLINE
PUBMED
REFERENCE

11
AUTHORS

12
TITLE

13
JOURNAL
MEDLINE
PUBMED
REFERENCE

14
AUTHORS

15
TITLE

16
JOURNAL
MEDLINE
PUBMED
REFERENCE

17
AUTHORS

18
TITLE

19
JOURNAL
MEDLINE
PUBMED
REFERENCE

20
AUTHORS

21
TITLE

22
JOURNAL
MEDLINE
PUBMED
REFERENCE

23
AUTHORS

24
TITLE

25
JOURNAL
MEDLINE
PUBMED
REFERENCE

26
AUTHORS

27
TITLE

28
JOURNAL
MEDLINE
PUBMED
REFERENCE

29
AUTHORS

30
TITLE

31
JOURNAL
MEDLINE
PUBMED
REFERENCE

32
AUTHORS

33
TITLE

34
JOURNAL
MEDLINE
PUBMED
REFERENCE

35
AUTHORS

36
TITLE

37
JOURNAL
MEDLINE
PUBMED
REFERENCE

38
AUTHORS

39
TITLE

40
JOURNAL
MEDLINE
PUBMED
REFERENCE

41
AUTHORS

42
TITLE

43
JOURNAL
MEDLINE
PUBMED
REFERENCE

44
AUTHORS

45
TITLE

46
JOURNAL
MEDLINE
PUBMED
REFERENCE

47
AUTHORS

48
TITLE

49
JOURNAL
MEDLINE
PUBMED
REFERENCE

50
AUTHORS

51
TITLE

52
JOURNAL
MEDLINE
PUBMED
REFERENCE

53
AUTHORS

54
TITLE

55
JOURNAL
MEDLINE
PUBMED
REFERENCE

56
AUTHORS

57
TITLE

58
JOURNAL
MEDLINE
PUBMED
REFERENCE

59
AUTHORS

60
TITLE

61
JOURNAL
MEDLINE
PUBMED
REFERENCE

62
AUTHORS

63
TITLE

64
JOURNAL
MEDLINE
PUBMED
REFERENCE

65
AUTHORS

66
TITLE

67
JOURNAL
MEDLINE
PUBMED
REFERENCE

68
AUTHORS

69
TITLE

70
JOURNAL
MEDLINE
PUBMED
REFERENCE

71
AUTHORS

72
TITLE

73
JOURNAL
MEDLINE
PUBMED
REFERENCE

74
AUTHORS

75
TITLE

76
JOURNAL
MEDLINE
PUBMED
REFERENCE

77
AUTHORS

78
TITLE

79
JOURNAL
MEDLINE
PUBMED
REFERENCE

80
AUTHORS

81
TITLE

82
JOURNAL
MEDLINE
PUBMED
REFERENCE

83
AUTHORS

84
TITLE

85
JOURNAL
MEDLINE
PUBMED
REFERENCE

86
AUTHORS

87
TITLE

88
JOURNAL
MEDLINE
PUBMED
REFERENCE

89
AUTHORS

90
TITLE

91
JOURNAL
MEDLINE
PUBMED
REFERENCE

92
AUTHORS

93
TITLE

94
JOURNAL
MEDLINE
PUBMED
REFERENCE

95
AUTHORS

96
TITLE

97
JOURNAL
MEDLINE
PUBMED
REFERENCE

98
AUTHORS

99
TITLE

100
JOURNAL
MEDLINE
PUBMED
REFERENCE

101
AUTHORS

102
TITLE

103
JOURNAL
MEDLINE
PUBMED
REFERENCE

104
AUTHORS

105
TITLE

106
JOURNAL
MEDLINE
PUBMED
REFERENCE

107
AUTHORS

108
TITLE

109
JOURNAL
MEDLINE
PUBMED
REFERENCE

110
AUTHORS

111
TITLE

112
JOURNAL
MEDLINE
PUBMED
REFERENCE

113
AUTHORS

114
TITLE

115
JOURNAL
MEDLINE
PUBMED
REFERENCE

116
AUTHORS

117
TITLE

118
JOURNAL
MEDLINE
PUBMED
REFERENCE

119
AUTHORS

120
TITLE

121
JOURNAL
MEDLINE
PUBMED
REFERENCE

122
AUTHORS

123
TITLE

124
JOURNAL
MEDLINE
PUBMED
REFERENCE

125
AUTHORS

126
TITLE

127
JOURNAL
MEDLINE
PUBMED
REFERENCE

128
AUTHORS

129
TITLE

130
JOURNAL
MEDLINE
PUBMED
REFERENCE

131
AUTHORS

132
TITLE

133
JOURNAL
MEDLINE
PUBMED
REFERENCE

134
AUTHORS

135
TITLE

136
JOURNAL
MEDLINE
PUBMED
REFERENCE

137
AUTHORS

138
TITLE

139
JOURNAL
MEDLINE
PUBMED
REFERENCE

140
AUTHORS

141
TITLE

142
JOURNAL
MEDLINE
PUBMED
REFERENCE

143
AUTHORS

144
TITLE

145
JOURNAL
MEDLINE
PUBMED
REFERENCE

146
AUTHORS

147
TITLE

148
JOURNAL
MEDLINE
PUBMED
REFERENCE

149
AUTHORS

150
TITLE

151
JOURNAL
MEDLINE
PUBMED
REFERENCE

152
AUTHORS

153
TITLE

154
JOURNAL
MEDLINE
PUBMED
REFERENCE

155
AUTHORS

156
TITLE

157
JOURNAL
MEDLINE
PUBMED
REFERENCE

158
AUTHORS

159
TITLE

160
JOURNAL
MEDLINE
PUBMED
REFERENCE

161
AUTHORS

162
TITLE

163
JOURNAL
MEDLINE
PUBMED
REFERENCE

164
AUTHORS

165
TITLE

166
JOURNAL
MEDLINE
PUBMED
REFERENCE

167
AUTHORS

168
TITLE

169
JOURNAL
MEDLINE
PUBMED
REFERENCE

170
AUTHORS

171
TITLE

172
JOURNAL
MEDLINE
PUBMED
REFERENCE

173
AUTHORS

174
TITLE

175
JOURNAL
MEDLINE
PUBMED
REFERENCE

176
AUTHORS

177
TITLE

178
JOURNAL
MEDLINE
PUBMED
REFERENCE

179
AUTHORS

180
TITLE

181
JOURNAL
MEDLINE
PUBMED
REFERENCE

182
AUTHORS

183
TITLE

184
JOURNAL
MEDLINE
PUBMED
REFERENCE

185
AUTHORS

186
TITLE

187
JOURNAL
MEDLINE
PUBMED
REFERENCE

188
AUTHORS

189
TITLE

190
JOURNAL
MEDLINE
PUBMED
REFERENCE

191
AUTHORS

192
TITLE

193
JOURNAL
MEDLINE
PUBMED
REFERENCE

194
AUTHORS

195
TITLE

196
JOURNAL
MEDLINE
PUBMED
REFERENCE

197
AUTHORS

198
TITLE

199
JOURNAL
MEDLINE
PUBMED
REFERENCE

200
AUTHORS

201
TITLE

202
JOURNAL
MEDLINE
PUBMED
REFERENCE

203
AUTHORS

204
TITLE

205
JOURNAL
MEDLINE
PUBMED
REFERENCE

206
AUTHORS

207
TITLE

208
JOURNAL
MEDLINE
PUBMED
REFERENCE

209
AUTHORS

210
TITLE

211
JOURNAL
MEDLINE
PUBMED
REFERENCE

212
AUTHORS

213
TITLE

214
JOURNAL
MEDLINE
PUBMED
REFERENCE

215
AUTHORS

216
TITLE

217
JOURNAL
MEDLINE
PUBMED
REFERENCE

218
AUTHORS

219
TITLE

220
JOURNAL
MEDLINE
PUBMED
REFERENCE

221
AUTHORS

222
TITLE

223
JOURNAL
MEDLINE
PUBMED
REFERENCE

224
AUTHORS

225
TITLE

226
JOURNAL
MEDLINE
PUBMED
REFERENCE

227
AUTHORS

228
TITLE

229
JOURNAL
MEDLINE
PUBMED
REFERENCE

230
AUTHORS

231
TITLE

232
JOURNAL
MEDLINE
PUBMED
REFERENCE

233
AUTHORS

234
TITLE

235
JOURNAL
MEDLINE
PUBMED
REFERENCE

236
AUTHORS

237
TITLE

238
JOURNAL
MEDLINE
PUBMED
REFERENCE

239
AUTHORS

240
TITLE

241
JOURNAL
MEDLINE
PUBMED
REFERENCE

242
AUTHORS

243
TITLE

244
JOURNAL
MEDLINE
PUBMED
REFERENCE

245
AUTHORS

246
TITLE

247
JOURNAL
MEDLINE
PUBMED
REFERENCE

248
AUTHORS

249
TITLE

250
JOURNAL
MEDLINE
PUBMED
REFERENCE

251
AUTHORS

252
TITLE

253
JOURNAL
MEDLINE
PUBMED
REFERENCE

254
AUTHORS

255
TITLE

256
JOURNAL
MEDLINE
PUBMED
REFERENCE

257
AUTHORS

258
TITLE

259
JOURNAL
MEDLINE
PUBMED
REFERENCE

260
AUTHORS

261
TITLE

262
JOURNAL
MEDLINE
PUBMED
REFERENCE

263
AUTHORS

264
TITLE

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAAGAGATCCAAAGACCTCTTTTCTTTTCTTTTCTTVA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAAGATTCGAGTTATTAATTAATTCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised longer than 7 kb was selected before cloning. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLc I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES

 o * Location/Qualifiers

 1. .3143

```

/clone.lib="RIKEN full-length enriched mouse cDNA library"
/dev.stage="adult"
327..3005
/note="unnamed protein product; DIPEPTIDYL PEPTIDASE 8
homolog [Homo sapiens] (SPR109HBK5, evidence: FASTA,
95.7%id, 100%length, match=2649)

```

```

/protein_id="BAB30295.2"
/db_xref="GI:26382128"
/db_xref="MG1:1921638"
/translation="MKIPSGRCNMMAAMETEOGVLEIFETACEGNGSDOPKLEIE
FYVERYSMSOLKTLADITRKHGYMMAKAHPDMVATGTPDRSHSRVYLITAGVG
RENTLFEYSIEPKTINRAAVLYLNMKPLDFOATLQDGMVSRREELIKERRIDIGI
AADYHPGSGTFLFOAGSGCYTHIKDGGPHETFOOPLRPNTVESCPIRLMDPKCPAD
PMDIATHSMDTIWISLYPHEPRITVYHNELNMEDPESACVATVLOEERPYGQ
YMMCPARBPSPGSKLRLIYKENDSESVETIHTVSMKTPPADSRPYRGTQANR
VYKMSRIYVUDNAGGILDYVDEKELVOPFEILFGVSEKTLAAGCMPECKAMSTLSDS
QHTQLVLSPELPIEVEDDAMRORLSEVSPSYPLIYIETDININHDITFVPI
QOHEBDEIERIPASCEKTPRHLYKLTSLIKESKYSKSGSLPAPSEPOKPIKAEIIT
TISEMEVLGRHSGINWDEARKUVEFGTKDLPHEHLVYTSANSEVRLVDTRGHS
HSCLSLHCDPLISKTSNQNPCHVSLYKSGSTPDEVEFKTEKFMALLDSAGLPPYR
TPLEISFESTTGFITGLMKPLKPHDQVQKSLPVLIVGQPOVQVNNKSPKCFYFR
LNTLASIGYVVVINDRGSCHRLKEFGAKYKMGQIEMDDQVEGYOYASLOPQDIL
DRTGHGMSYGYVLSLMALMGQRSDIYRATAGAPVAILIYDGYETRYVNGHPDNDL
GYVLGYSVAMQAEKTPSEPNNRLIHLGFLDENVEHFATSLISLFLVRAGKPYDLOIYPO
ERSIRVPEGEHVELLHLVLOENIGSRIALAKVI"

```

Alignment Scores:	
Pred. No.:	1,23e-190
Score:	1756.50
Percent Similarity:	65.67%
Best Local Similarity:	64.33%
Query Match:	90.87%
DB:	11
	Gaps:
	1

QY 1 GluCluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluPheAspArg 20
US-10-070-464-7 (1-360) X AK016546 (1-3143)

Db 1086 GAAAGAGATCCAGATCAAGCTGATGTGGCACCTTTGTTCTTCAAGAAATTTTGACAA 1145
 Qy 21 TysSerGlyTyrTrpCysProLysAlaGluThrProSerGlyGlyLysIleLeu 40
 Db 1146 TACTCTGGCTACTGGTGGTGTCCCAAGGGAAGAACTCTAGTGTGTAATAATCTT 1205
 Qy 41 ArgGlyLeuTyrGluGlnAsnAspGluSerGluValGluIleLeuHisValThrSerPro 60
 Db 1206 AAGAAATCTCTATGAAGAAATATGATGATCTGAGGTGAGATTAATTCATGTCCTCC 1265
 Qy 61 MetLeuGluThrArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
 Db 1266 ATGTTGGAAACAAGAGGCAATTCCTTCGTTATCCAAAACAGGCAACCAACCA 1325
 Qy 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle 98
 Db 1326 AAGGTCACTTCAAGATGTGGAAGATTGTTGATGCTGCGAGAGGAGATTATAGATGC 1385
 Qy 98 ----- 98
 Db 1386 ATGATATAAGAACTGGATTCMACTTTCGAGATTCTGTTGAGGAGTTGATATATATGCC 1445
 Qy 98 ----- 98
 Db 1446 AAGGCGGATGATCTCCAGAGGAAACATGCTGATCTCATCTAGACCGTTCCAG 1505
 Qy 98 ----- 98
 Db 1506 ACTCACTGCAGATAGTCTGTATCTCCCTGAGTTATTCATCCAGTAAAGATGATGCC 1565
 Qy 98 ----- 98
 Db 1566 ATGACAGACAGAGACTTATAGAGTCAATTCTGACTGTGACACACTGATCATCTAT 1625
 Qy 98 ----- 98
 Db 1626 GAAGAAACAACAGACATCTGATAATAATCCACGATATTTTTCATGTTTCTCTCAACT 1685
 Qy 98 ----- 98
 Db 1686 CATGAAGATGAATGAGTTATTTTGGCTCTGATGACAAACAGGTTTGTGATCTG 1745
 Qy 98 ----- 98
 Db 1746 TATAAATCAATCCATTTTAAAGAGAGCAATATAACGGTCCAGTGTGACTACT 1805
 Qy 98 ----- 98
 Db 1806 GCCCCAAGTATTTCAAGTGTCTTATCAAAAGAAATATACATTACAGTGTGAATGG 1865
 Qy 99 -----ValAspGluValArgArgLeuValTyr 107
 Db 1866 GAAGTACTTGGCCGGCATGATCTAATATCTGGTTGATGAAGCCAGAAAGCTGTGTAC 1925
 Qy 108 PheGluGlyThrLysAspSerProLeuGlnHisIleuTyrValValSerTyrValAsn 127
 Db 1926 TTGTAAGGCAACCAAGACTCTCTTTGAGCACTCACTGTAGTACCAAGTATATCAAC 1985
 Qy 128 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
 Db 1986 CCTGAGAAAGTGTGAGGCTGATCTGACCGTGTCTACTCACTCTGCTGCTCCGCCGG 2045
 Qy 148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
 Db 2046 CATGTGTACTTCTCATATAGTATAGTACAGCAACAGAAAGATCCACACGTGTGTCTC 2105
 Qy 168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGlnPheThrAlaThr 187
 Db 2106 TACAAACTCTCAAGTCTGAGAGATGACCAAGTTCATTAACCAAGAAATTTTGGCCACC 2165
 Qy 188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 207

Db 2166 ATTTGGATTACGACAGTCTTCTTCTGACTACACCCCTCCAGAAATTTTCTTTGAA 2225
 Qy 208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPheHisAspLeuGlnProGly 227
 Db 2226 AGTACTACTGATGATTTACACTGATATGGAATGTTGATTAAGCCATGACCTACACCTGA 2285
 Qy 228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
 Db 2286 AAGAAATACCCCACTGTGTTATTCATATATGTGTGTGCCCAAGTGCAGCTGTGGAACAAT 2345
 Qy 248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
 Db 2346 CGGTTTAAAGAGTCAAGATTTCCGCTGAACACCTGCGCTCTCTGAGTTATGTGGTT 2405
 Qy 268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
 Db 2406 GTGGTGATAGACAAACAGGGGATCTCTGTACCGAGACCTTAATTTGAAGCCCTTTAA 2465
 Qy 288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
 Db 2466 TATAAAATGGGTCAATATGAATCGATGATCAAGTGAAGGACTCCAGTACCTAGCATCT 2525
 Qy 308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyThrSerTyrGlyTyr 327
 Db 2526 CAGTATGACTTATGACTGATGATGAGTGGCATCCACGAGACTTAATTTGAAGCCCTTAA 2585
 Qy 328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
 Db 2586 CTCTGCTGATGATGATTAATGACAGAGTGGATATCTTCGGGTGCTATGTGTGGGCC 2645
 Qy 348 ProValThrLeuThrIlePheTyrAspThrGlyTyrThr 360
 Db 2646 CCAGTACACCTGTGTGATCTTATGATACAGATACAG 2684
 RESULT 2
 AK029788
 LOCUS
 DEFINITION
 Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930560C15 Product:DIPEPTIDYL PEPTIDASE 8 homolog [Homo sapiens], full insert sequence.
 AK029788.1 GI:26081520
 VERSION
 AK029788.1
 KEYWORDS
 HTG; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 JOURNAL
 99279253
 PUBMED
 10349636
 AUTHORS
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 JOURNAL
 20499374
 MEDLINE
 11042159
 PUBMED
 11042159
 REFERENCES
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K., Fujiwaga, S., Inoue, K., Togawa, Y., Iawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL
 20530913
 MEDLINE

PUBMED
REFERENCE
AUTHORS
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Komono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kado, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudli, F., Suzuki, R., Tomita, M., Wagner, J., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carrinci, P., de Bona, M. P., Brownstein, M. T., Bulc, C., Hill, D., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.

TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
REFERENCE
AUTHORS
5
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 5517)

TITLE
JOURNAL
REFERENCE
AUTHORS
6
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carrinci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 5517
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4930560C15"
/db_xref="taxon:10090"
/clone="4930560C15"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
2221. 4837

FEATURES
source
misc_feature

/note="DIPPEPTIDYL PEPTIDASE 8 homolog (Homo sapiens) (SPTR|Q9HBM5, evidence: FASTA, 95.7%ID, 100%length, match=2649) putative"
BASE COUNT 1514 a 1224 c 1230 g 1549 t
ORIGIN
Alignment Scores: 3,556-180 Length: 5517
Pred. No.: 1669.00 Matches: 332
Score: 66.67% Conservative: 10
Percent Similarity: 64.72% Mismatches: 18
Best Local Similarity: 86.34% Indels: 133
Query Match: 11 Gaps: 4
DB: 11
US-10-070-464-7 (1-360) x AK029788 (1-5517)
QY 1 GUGUUAAPALAAxSerIlaGlyValAlaThrPheValLeuGluGluGluPheAspArg 20
|||
Db 2980 GAAGAGGATCCCAAGATCAAGCTGCTGTCACCTTGTCTTCAAGAAATTGACAGA 3039
|||
QY 21 TYSerGlyTyrTrpTyrCysProIysAlaGluThrThrProSerGlyGlyIleLeu 40
|||
Db 3040 TACTCTGGCTACTGGTGGTGGTCCCAAGCGAAGAACTCCTAGTGTGTAATTTCTT 3099
|||
QY 41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleValThrSerPro 60
|||
Db 3100 AGAATTCCTATGAAGAAATATGATGATGAGGAGGATTAATCATGATCGTCCCC 3159
|||
QY 61 MetLeuGluThrArgArgIlaAspSerPheArgTyrProIys----- 74
|||
Db 3160 ATGTTGAAACAAAGAGGGGCAAGATTCCTTGGTATCCAAACAGATTATGATGTCA 3219
|||
QY 75 -----ThGlyThrAla----- 78
|||
Db 3220 AGATAAGAACTGTTCAACCTTGAGATCTGTTGAGGAGTTGAATATATTGCCAG 3279
|||
QY 78 ----- 78
|||
Db 3280 AGCCGGATGACTCCAGAGGAAACATCGTTCATCTACTAGACCGTCCAGAC 3339
|||
QY 79 AsnPro-LysValThrPheIleMetSerGluIleMetIleAspAlaGlu----- 94
|||
Db 3340 TCACCTGCAATAGTTCTGATCTCCCTGATTAATTCACAGTAGAAGATGATGCCAT 3399
|||
QY 94 ----- 94
|||
Db 3400 GGACAGACAGAGACTTATAGAGTCAAGTCTCGACTGTGACACACATCATCTATGA 3459
|||
QY 94 ----- 94
|||
Db 3460 AGAAACAGACAGACATGTGATAAATATCCAGATATTTTCATGTTTTCCTCAACTCA 3519
|||
QY 94 ----- 94
|||
Db 3520 TGAAGATGAATTGAGTTATTTTTCCTGCTGAATGCCAAACAGTTTCGTATCTGTA 3579
|||
QY 94 ----- 94
|||
Db 3580 TAAATCAATCCATTTTAAAGAGAGCAATATAACGTCGACGTGTGACTACTGC 3639
|||
QY 94 ----- 94
|||
Db 3640 CCCAAGTATTTCAGATGTCCTATCAAGAGAAATTAACAATTACCAAGTGTGAATGGGA 3699
|||
QY 95 -----GlyArg-----IleIleValAspGluValArgSerValTyrPhe 108
|||
Db 3700 AGAAGTGGCCCGGCAATGATTAATATCGGGTGTGTAAGCCAGAAAGCTGTGATCTT 3759
|||
QY 108 eGluGlyThrLysAspSerProLeuGluIleHisIleValThrValSerTyrValAsnPr 128
|||
Db 3760 TGAAGCACCAAAAGACTCTTCTTGGAGCATCACTGTGACGTAAGCAAGTTATGCAACC 3819
|||

QY	128	IGLJ1G1Val1ThirArgLeuThrAspArgGlyTYRSeH1SerCysCys111SerGlnH1	148
Db	3820	TGGAGAAAGTGTGAGGCTGACTGACCGTGGCTACTCACTCCGCTCCTCAAGCCGCA	3879
QY	148	SCysAspPheH1eSerLySTYRSeAsnGlnLysAsnProH1sCysVal1SerLeuTY	168
Db	3880	TTGTGACTTCTTCATTAAGTAAGTACAGCAACCAAGAAATCCACTGTGTGCTCCTTA	3939
QY	168	RLysLeuSerSerProGluAspAspProThrCysLysThirLysGlu1UpH1eTPA1aThr11	188
Db	3940	CAAACTCTCAAGCTCTTGAGATGACCCAGTTCATATAAACAAAGAAATTTTGGCCACAT	3999
QY	188	eLeuAspSer1aGlyProLeuProAspTYRThrProProGlu11ePheSerPheGluSe	208
Db	4000	TTTGATATTACACAGGTCTCTCTCGATACACCCCTCCAGAAATTTTCTTTTGAAG	4055
QY	208	rThThrGlyPheThrLeuTYRGLYMetLeuTYR1LysProH1sAspLeuGlnProGlyLY	228
Db	4060	TACTACTGATTTACACTGATGGAATGTTGATATAGCCCTCATGACCTACCAACCTGGAA	4119
QY	228	sLySTYRProThrVal1LeuPhe11eTYRGLY1ProGlnVal1GlnLeuValAsnAspAr	248
Db	4120	GAAATACCCCACTGTGTTATCATATATAGTGATGCCCAAGG1GACGCTGGTGAACATCG	4179
QY	248	gPheLysG1Val1LysTYRPhaArgLysAsnThrLeuVal1aSerLeuG1TYRVal1Val1a	268
Db	4180	GTTTAAAGAGATCAAGTATTTCCGCTGAAACACCTGGCTCCTCCGTATGTATGTGTGT	4239
QY	268	lVal11aAspAsnArgGlySerCysH1sArgGlyLeuLysPheGluGlyVal1aPheLysTY	288
Db	4240	GGTATATGACACACAGGGATCTCTGCACCGAGGATTTAAATTTGAAGGCGCTTTAAATA	4299
QY	288	rLYMetG1Y61n11eGlu11eAspAspG1NVal1Glu1LysLeuGlnTYRLeuVal1aSerAr	308
Db	4300	TAAATGGGTCAATTAAGAAATCGATGATCAAGTGGAAAGACTCCAGTACCTAGCATCTCA	4355
QY	308	gTYRAspPhe11eAspLeuAspArgVal1Gly11eH1sGlyTrpSerTYRGLYTYRLe	328
Db	4360	GTATGACTTCATTACTGATCTTGATGCGATGCGGATCCACCGCTGGTCTCTTATGTGTGCTACCT	4419
QY	328	uSerLeuMet1aLeuMetG1NArgSerAsp11ePheArgVal1a11e1aGlyVal1aPr	348
Db	4420	CTCCCTGATGSCATTAAATGACAGAGGTCCGAAATCTTCCGGGTGGCTATTGCTGGGCCCC	4479
QY	348	oVal1ThrLeuTrp11ePheTYRAspThrGlyTYRThr	360
Db	4480	AGTCAACCTGTGATCTCTATGATACAGATACAG	4516
RESULT 3			
BM557438			
LOCUS			
DEFINITION	AGNCOCURT 6578992 NIH_MGC_41 Homo sapiens cDNA IMAGE:5466759		
ACCESSION	5', mRNA sequence.		
KEYWORDS	BM557438		
SOURCE	BM557438.1 GI:18799430		
ORGANISM	EST.		
	Homo sapiens (human)		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 1042)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: DCTD/DTF cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.nih.gov		

FEATURES	source	1..1042
Plate: L1CM1967 row: k column: 16		
High quality sequence stop: 697.		
Location/Qualifiers		
		1..1042
		/organism="Homo sapiens"
		/mol_type="rRNA"
		/db_xref="taxon:9606"
		/clone="IMAGE:5466759"
		/tissue_type="amelanotic melanoma, cell line"
		/lab_host="DH10B (phage-resistant)"
		/clone_id="NH MGC 41"
		/note="Organ: skin; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."
BASE COUNT	279 a	223 c 234 g 303 t 3 others
ORIGIN		
Alignment Scores:		
Pred. No.:	4.97e-152	Length: 1042
Score:	1415.00	Matches: 262
Percent Similarity:	99.24%	Conservative: 0
Best Local Similarity:	99.24%	Mismatches: 2
Query Match:	73.20%	Indels: 0
Ds:	12	Gaps: 0
US-10-070-464-7 (1-360) x BM557438 (1-1042)		
Qy	97	11e11eVaAspgluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 116
Ds	28	ATCCAGTTGATGANGTCAGAGGCTGGTATATTTTGAAGCACCAAGACTCCCTTTA 87
Qy	117	GLuHsHsLeuTyValValSerTyValAsnProGlyGluValThrArgLeuThrAsp 136
Ds	88	GAGCATCACTGTAGTCAGTCACTAGTACGTAATCTCGAGAGGTGACCAAGCTGACTGAC 147
Qy	137	ArgGlyTyrSerHisSerCysCysAlaSerGlnHisCysAspPheHelleSerLysTyr 156
Ds	148	CGTGCTCACTACATCTTCTGCTGATCGATCGACCTGACTGACTCTTTTAACTAAAGTAT 207
Qy	157	SerAsnGlnLysAsnProHisCysValSerLeuTyLysLeuSerSerProGluAspAsp 176
Ds	208	AGTAAACCGAAGAAATCCACACTGTGTGTCCCTTCAAGCTATCAAGTCTGAAGATGAC 267
Qy	177	ProThrCysLysThrLysGluPheThrAlaThrLleLeuAspSerAlaGlyProLeuPro 196
Ds	268	CCAACTTGCAAAACAAAGAAATTTGGGCGCACCATTTGGATTGACTGACAGGCTCTTCCCT 327
Qy	197	AspTyrThrProProGlnLlePheSerPheGluSerThrThrGlyPheThrLeuTyRgly 216
Ds	328	GACTATACTCTCCAGAAATTTCTCTTTGAAGTACTACTGAGATTTCATTGATGAGG 387
Qy	237	MetLeuTyLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 236
Ds	388	ATGCTCTCAAGACCTCATGATCTACAGCTGGAAAGAAATATCTACTGTGCGTTCCATA 447
Qy	237	TyrGlyGlyProGlnValGlnLeuValAsnAsnAsgPheLysGlyValLysTyrPheArg 256
Ds	448	TATGGTGTCTCCAGAGTCCAGTGTGATATATCGTTTAAAGGAGTCAAGTATTTCCGC 507
Qy	257	LeuAsnThrLeuAlaSerLeuGlyTyrValValValLleAspAsnArgLysSerCys 276
Ds	508	TTGATATACCTTACGCTCTCTAGGTTATGTGTGTATGATGATGACACACAGGGATCTGT 567
Qy	277	HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluLleAsp 286
Ds	568	CACCGAGGCTTAAATTTGAAGCGCCCTTTTAAATTTAAATGGCTCAATATAGAAATTTGAC 627
Qy	297	AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheLleAspLeuAspArg 316

Db 628 GATCAGGTGAGAGACTCCAAATATCTAGCTTCTCGATATGATTTCTTACCTTAGATCGT 687
 QY 317 VALGILYIIEHISGLYTPserTYrGLyTYrLeuSerLeuMetAlaIeuMetGlnArg 336
 Db 688 GTGGGATCATCAGCGCTGTCCTATGAGAGATACCTCTCCCTGATGSCATTATGACGAG 747
 QY 337 SerASpIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTYrAsp 356
 Db 748 TAGAGATCTTCAGGGCTTCATCTGCTGGGGCCCGACCTCTGTGATCTTCTATGAT 807
 QY 357 ThrGLYTYrThr 360
 Db 808 ACGAGTACACG 819
 RESULT 4
 LOCUS BQ675260 910 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT 8354972 NIH_MGC_102 Homo sapiens CDNA clone IMAGE:6275478
 ACCESSION BQ675260
 VERSION BQ675260.1 GI:21786094
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 910)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LIML)
 Clone distribution by: Agencourt Bioscience Corporation
 found through the I.M.A.G.E. Consortium/LIML at:
 http://image.llnl.gov
 Plate: L1CM2458 row: 1 column: 07
 High quality sequence stop: 618.
 Location/Qualifiers
 source 1..910
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6275478"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 102"
 /note="Organ: salivary gland; Vector: pOT81; Site_1: XhoI;
 Site_2: EcoRI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGAG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-CDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies). Note: this is a NIH-MGC Library."
 BASE COUNT 248 a 186 c 212 g 262 t 2 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 8, 28e-142 Length: 910
 Score: 1325.50 Matches: 254
 Percent Similarity: 96.24% Conservative: 2
 Best Local Similarity: 95.49% Mismatches: 6
 Query Match: 68, 57% Indels: 4
 DB: 13 Gaps: 1
 US-10-070-464-7 (1-360) X BQ675260 (1-910)
 QY 97 IIEIIEVALASpGluValArgArgLeuValTYrPheGluGlyThryLysAspSerProLeu 116

Db 116 ATCCAGATTGAGAGAGTCAAGAGCTGATATTTTGAAGACCAAAACCTCCCTTTA 175
 QY 117 GIuHISLLeuTYrValValSerTYrValAsnProGlyGluValThrArgLeuThrAsp 136
 Db 176 GAGCATCACTCTAGCTAGTACGTTACGTAATCTCTGAGAGAGTGAACAAGCTGACTGAC 235
 QY 137 ArgGLYTYrSerHisSerCysGlyIleSerGlnHisCysAspPhePheIleSerLYr 156
 Db 236 CGTGGCTACTCAATCTTGCTGACATCACTGACACCTGACCTCTTTTAAGTAGAT 295
 QY 157 SerAsnGlnLysAsnProHisCysValSerLeuTYrLysLeuSerSerProGluAspAsp 176
 Db 296 AGTAACCAAGAAATCCACACTGTGTCTCTTACCACTATCAAGTCTCAAGATGAC 355
 QY 177 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 196
 Db 356 CCAACTGGCAAAACAAGAAATTTGGGCGACACATTTGGATTACACAGCTCTCTTCCT 415
 QY 197 AspTYrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTYrGly 216
 Db 416 GACTATACCTCTCCAGAAATTTCTCTTTGAAAGTACTGATTTACATTTGATGGG 475
 QY 217 MetLeuTYrLysProHisAspLeuGlnProGlyLYrLysLYrProThrValLeuPheIle 236
 Db 476 ATGCTCTACAAGCTCATGATCTACAGCTGGAAGAAATATCTACTGCTGCTTCATA 535
 QY 237 TYrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLYrTYrPheArg 256
 Db 536 TAGGGGTCCTCAGGTGAGTGGATTAATCCGTTTAAAGAGACTCAAGTATTTCCGC 595
 QY 257 LeuAspThrLeuAlaSerLeuGlyTYrValValValIleLysAsnArgLysSerCys 276
 Db 596 TTGAATACCTTAGCCTCTTAGATGATGTGTGATGATGACAAACGGGATCTGT 655
 QY 277 HISATGlyLeuLysPheGluGlyAlaPheLysTYrLysMetGlyGlnIleGluIleAsp 296
 Db 656 CACCGAGGGCTTAATTAATTAAGAGCGCTTTAATATTAATGATGCTAATAGAAATGAC 715
 QY 297 AspGlnValGluGlyLeuGlnTYrLeuAlaSerArgTYrAspPheIleAspLeuAspArg 316
 Db 716 GATCAGGTGAGAGAGCTCCAAATATCTAGCTTCTCGATATGATTTCACTTAGATCGT 775
 QY 317 VALGILYIIEHISGLYTPserTYrGLyTYrLeuSerLeuMetAlaIeuMetGlnArg 336
 Db 776 GTGGGATCATCAGCGCTGTCCTATGAGAGATACCTCTCCCTGATGSCATTATGACGAG 835
 QY 336 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrp---Ile-Phe 354
 Db 836 GTCAAGATTATCTTCAAGTTGCTATGCTGGGGCCCGACANTCAATCTGGGGGATCTTTC 895
 QY 355 TYrAspThrGly 358
 Db 896 TATGATACAGGG 907
 RESULT 5
 LOCUS BQ675006 957 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT 8303592 NIH_MGC_102 Homo sapiens CDNA clone IMAGE:6274622
 ACCESSION BQ675006
 VERSION BQ675006.1 GI:21785840
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 957)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LNCM2456 row: h column: 15
 High quality sequence start: 4
 High quality sequence stop: 624.
 Location/Qualifiers

FEATURES

1..957
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5274622"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_102"
 /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-cDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies). Note: this is a NIH_MGC Library."
 BASE COUNT 257 a 206 c 228 g 266 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.72e-138 Length: 957
 Score: 1297.00 Matches: 253
 Percent Similarity: 95.15% Conservative: 2
 Best Local Similarity: 94.40% Mismatches: 8
 Query Match: 67.10% Indels: 5
 DB: 13 Gaps: 0

US-10-070-464-7 (1-360) x BQ675006 (1-957)

QY 97 lleleValaSPglValARgArleuValTYrPhelgluTYrThrysaSPserProleu 116
 Db 129 ATCCAAgTtGATGAAgttCAGAAgGCTGtATATTTTGAAGGACCAAGACTCCCTTTA 188
 QY 117 GluHSHSLeuTYrValaSerTYrValaSProglYgluValThraArgleuThraSP 136
 Db 189 GAGCATCACCTGACTAGTACAGTAAATCTCGAAGGTGACAAAGCTGATGAC 248
 QY 137 ArgGlyTYrSerHisSerCysIleSerGlnHisCysAspPhehelIleSerlyTYr 156
 Db 249 CCGTGCTACTCATTCTTGCTGCACAGTCAGCAGCTGACTCTTTATTAAGTAAGTAT 308
 QY 157 SerAANGlnLYaSPProHIScysValaSerleuTYrLYaSPserProgluAspAP 176
 Db 309 AGTAACCAAGAAATCAACACTGTGTCTCTTACAAAGCTATCAAGTCTGAAAGTAC 368
 QY 177 ProthrCysLYaSPThrLYaSPThrPAlaThrIleLeuAspSerlaGlyProleuPro 196
 Db 369 CCAACTGCAGAAACAAAGAAATTTTGGCCACATTGGATTGAGTACAGGTCTCTCTCT 428
 QY 197 AspyTYrThrProProgluIlePheSerPhegluSerThrThglYpHeThreutyTYr 216
 Db 429 GACTATACCTCCCGAAATTTCTCTTTGAAAGTACATCGGATTACATTGATGAGG 488
 QY 217 MetLeuTYrLYaSPProHISaSPleuGlnProglYLYaSPThrProthrValaLeuPhe 236
 Db 489 ATGCTCTACAGCTCATATCTACAGCTCGAAGAAATATCTTCTGTGCTTTCA 548
 QY 237 TYrGlyLYaSPProGlnValaGlnLeuValaAsnaSPArgPheLYaSPValaLYaSP 256
 Db 549 TATGCTGTGCTCAGGTGAGTGAATATCGGTTAAAGAGACTCAAGTATTTCCGC 608
 QY 257 LeuAenThr-LeuAlaSerleuGlyTYrValaValaValaIleAsp-AsnaArgGlySerC 276

Db 609 TTGAATACCCCTTACGCTCTTATGTTATGTGTATGATGATGACAAACAGGATCCT 668
 QY 276 YHIArGglYleuLYaSPThrPhegluGlyAla-PheLYaSPMet-GlyGlnIleGluI 295
 Db 669 GTACACGAGGCTTAAATTTGAAGGCGCCCTTTATATATAAATGGGTCMAATAGAAAT 728
 QY 285 eAspAspGlnValaGluLYaSPGlnTYrleuAlaSerArgTYrAspPheIleAspPhe 315
 Db 729 TCACGATGAGGGGAGAGAGCTCCAAATATCTAGCTTCTCATATGATTTATTGATTA 788
 QY 315 PaTYrValaGlyIleHISGlyTYrPserTYrGlyGlyTYrLeuSerleuMetAlaLeuMetG 335
 Db 789 TCGTGTGGGCATCCACGCTGCTCATATGAGAGATCCCTCCCTGAAGGCAATTATGCA 848
 QY 335 nArgSerAspIlePheArgValaAlaIleAlaGlyAlaSProValThr-LeuThrIlePheT 355
 Db 849 GAGGTAGATATCTTCAGATGTCTATGTCTGGGGGCCAGTACCTCGTGGATCTTCC 908
 QY 355 yRAspThrGlyTYr 359
 Db 909 ATGATACAGATAC 922

RESULT 6
 BI084090 1041 bp mRNA linear EST 20-JUN-2001
 LOCUS 602869453P1 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:5013996 5',
 DEFINITION mRNA sequence.
 BI084090
 ACCESSION BI084090.1 GI:14502420
 VERSION
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1041)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contract: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LNCM1820 row: j column: 13
 High quality sequence stop: 847.
 Location/Qualifiers

FEATURES

1..1041
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5013996"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_102"
 /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-cDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies). Note: this is a NIH_MGC Library."
 BASE COUNT 280 a 227 c 242 g 292 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.41e-131 Length: 1041
 Score: 1237.50 Matches: 254

Percent Similarity: 93.77%
 Best Local Similarity: 93.04%
 Query Match: 64.02%
 DB: 12
 Gaps: 0

US-10-070-464-7 (1-360) x BI084090 (1-1041)

QY 97 IIEIleValAspGluValArgArgLeuValTyrPheGluGlyThrTyrAspSerProLeu 116
 DB 120 ATCCAGTGTGATGAACTCAGAAAGCTGCGATATATTTGAAAGGCGCCAAAGACTCCCTTTA 179
 QY 117 GIuHSHleuTyrTyrValValSerTyrValAspProGluGluValThrArgLeuThrAsp 136
 DB 180 GAGCATCACCCTGTTAGCTAGTCACTTCGTAATCTCTGAGAGAGGTGACAGGCTGACTGAC 239
 QY 137 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyr 156
 DB 240 CGTGCTACTCATTCTTGCTGTCATCAGTCACTGAGCTGAGCTCTTTATAGTAACTAT 299
 QY 157 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 176
 DB 300 AGTAACCAAGATCCACACTGTGTGCTCCCTTTACAGCTATCAAGTCTCTGAAAGATGAC 359
 QY 177 ProThrCysLysThrLysGluPheTyrPalaThrIleLeuAspSerAlaGlyProLeuPro 196
 DB 360 CCACCTTGCAGAAACAAAGAAATTTGGGCCACATTTGGATTCAAGAGTCTCTTCT 419
 QY 197 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 216
 DB 420 GACTAATACCTCCCAAAATTTCTCTTTGAAAGTACTACATTTGATTTGATTTGAGG 479
 QY 217 MetLeuTyrLysProHisAspLeuGlnProGlyLysLeuTyrProThrValLeuPheIle 236
 DB 480 ATGCTCTACAGGCTCATGATCTACAGCTCGAAAGAAATCTTACTGTGCTGTTCA 539
 QY 237 TTYGIGlyProGlnValGlnLeuValAsnAsnArgPheIleGly--ValIleTyrPheA 256
 DB 540 TATGGTGGTCTCCTAGGTGAGTTCGTAATACGCTTTAAACGCAAGTCAATATTTCC 599
 QY 256 rgleuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerC 276
 DB 600 GCTTGAATACCCAGCTCTCTGATTTAGTGTGTTAGTATGATGAGCAACAGGGGATCC 659
 QY 276 YSHISrGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleA 296
 DB 660 GTACCGAGGCTTAATTTGAAAGGCGCTTTAA-TATACATGGGTCAATA-GAATTTG 717
 QY 296 spAspGlnValGluGlyLeuGln-TyrLeuAlaSerArgTyrAspPheIle-AspLeuAs 315
 DB 718 AGCATCAGCTGGAAGGACTCCCAATTAATCTAGCTTCTCGAATGATTTCACTTTGACTTGA 777
 QY 315 PATGValGlyIleHisGlyTyrSerTyrGly-GlyTyrLeuSerLeu-MetAlaLeuMet 334
 DB 778 TCGTGGGCGATCCAGGCGGTGTCATAGGAGGAACTCTCCCGAATGGCATTAAG 837
 QY 335 GlnArgSerAsp-IlePheArgValAlaIleAla-GlyAlaProValThrLeuTyrIleP 354
 DB 838 CAGAGTTCAGATCTCTTCCAGGCTCATAGCTGCGGGGCCCAAGTCACTCTGTGATCT 897

QY 354 he-TyrAspThrGlyTyrThr 360
 DB 898 TCTTATGATACAGGATACAG 918

RESULT 7
 BY751026 746 bp mRNA linear EST 17-DEC-2002
 LOCUS BY751026
 DEFINITION CDNA clone F830018D07 5', mRNA sequence.
 ACCESSION BY751026
 VERSION BY751026.1 GI:27181193
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 746)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Oosato, N., Saito, R., Suzuki, H., Yamana, K., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojodori, I., Balderelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Knapkin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusic, V., Chothia, C., Corbett, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, J. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzia, L., Miki, H., Nagashima, T., Nmatra, K., Okido, T., Pavan, W. J., Pettes, G., Pesole, G., Petrowsky, N., Pillai, R., Pontus, J. U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semp, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wymshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arikawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

TITLE JOURNAL MEDLINE PUBMED COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenihiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, U., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Oosato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Watanabe, M. and Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 eComputer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

source

Location/Qualifiers

1..746

/organism="Mus musculus"

/mol_type="mRNA"

/strain="NOD"

/db_xref="taxon:10090"

/clone="F830018D07"

/issue_type="activated spleen"

/clone_lib="RIKEN full-length enriched, activated spleen"

BASE COUNT 194 a 178 c 176 g 197 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 3.62e-131 Length: 746

Score: 1232.00 Matches: 229

Percent Similarity: 97.48% Conservative: 3

Best Local Similarity: 96.22% Mismatches: 5

Query Match: 63.74% Indels: 1

DB: 14 Gaps: 0

US-10-070-464-7 (1-360) x BY751026 (1-746)

QY 124 SerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCys 143

DB 1 AGTTATGCAAAACCTGAGAGAGTGTGAGGCTGACTGACCGTGGCTACTGACCTCTGC 60

QY 144 CysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnAspProHis 163

DB 61 TGGCTACGCGGATGTGACTTCTTCATAGTAAGTACGCAACCAAGAAATCCACAC 120

QY 164 CysValSerLeuTyrIleuSerSerProGluAspAspProThrCysIleTyrGlu 183

DB 121 TGTGTGTCCTCTACCAACTCTCAAGTCCGAGAGATGACCCAGTTCTATAAACAAAGAA 180

QY 184 PheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspPheThrProProGluIle 203

DB 181 TTGGGACCAACCTTGGATGACAGGTCCTCTTCGACTACACCCCTCAGAAAT 240

QY 204 PheSerPheGluSerThrThrGlyPheThrIleuTyrGlyMetLeuTyrIleuProHisAsp 223

DB 241 TTTCTTTGMAAGTACTGACTGATTTACCTGTATGGAATGTGTATTAAGCTCATGAC 300

QY 224 LeuGlnProGlyIleuSerTyrProThrValLeuPheIleTyrGlyGlyProGlnValGln 243

DB 301 CTACAACTCGAAGAAATACCCCACTGTTATATATATGTTGTTCCCGAGTGAC 360

QY 244 LeuValAspAspArgPheLeuGlyValIleTyrPheArgPheAsnThrIleuAlaSerIleu 263

DB 361 CTGGTGAACATCGGTTTAAAGAGTCAAGTATTTCCGCTGAAACCCCTGGCTCCCTG 420

QY 264 GlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuIleuSerPheGlu 283

DB 421 GGTATGTGTTGTGGATAGACAAAGGAGTCCGTGACCGAGGACTTAATTTGAA 480

QY 284 GlyAlaPheLeuTyrIleuSerGlyGlnIleGlnIleAspAspGlnValGlnIleLeuGln 303

DB 481 GGGGCTTTAAATATATAATGGGTCAATATGAATGATGATCAATGTGAAGACCTCCAG 540

QY 304 TyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSer 323

DB 541 TACCTAGCATCTCAGATGATGATCTTCACTTGGATCGAGTGGGCAATCCAGGCTGATCC 600

QY 324 TyrGlyGlyTyrIleuSerLeuMetAlaLeuMetGlnArgSerAspIle-PheArgValAla 343

DB 601 TATGGTGGTACTCTCCCTGATGGATTAATGCAAGGTCGATATCTGTCGGGGTGGG 660

QY 343 AlaLeuAlaGlyAlaProValThrLeuThrIlePhePheTyrAspThrGlyTyrThr 360

DB 661 TATTGTGGGGCCCGCAGTCACTGTGGATCTTCTATGATACAGATACAG 712

RESULT 8

BO671635

LOCUS BO671635 968 bp mRNA linear EST 15-JUL-2002

DEFINITION

AGENCOURT 8203887 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255645

5', mRNA sequence.

ACCESSION

BO671635

VERSION

BO671635.1 GI:21782469

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 968)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

CONTACT: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

plate: LCM2407 row: a column: 22

High quality sequence stop: 630.

Location/Qualifiers

1..968

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6255645"

/issue_type="epidermoid carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 102"

/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCCGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

BASE COUNT 245 a 228 c 235 g 259 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 7e-128 Length: 968

Score: 1205.00 Matches: 229

Percent Similarity: 97.46% Conservative: 1

Best Local Similarity: 97.03% Mismatches: 3

Query Match: 62.34% Indels: 3

DB: 13 Gaps: 0

US-10-070-464-7 (1-360) x BO671635 (1-968)

QY 97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrIleAspSerProLeu 116

DB 117 ATCCAAGTGAAGTGAAGTCAAGGCTGTAATATTTGAAGGCACCAAGACTCCCTTTA 176

QY 117 GluHisIleLeuTyrAlaValSerTyrValAspProGlyGluValThrArgLeuThrAsp 136

DB 177 GAGCATCACTGATCACTGATCACTTACGTAATCCGGAAGGTGACAAAGGCTGACTGAC 226

QY 137 ArgGlyTyrSerHisSerCysIleSerGlnHisCysAspPhePheIleSerIleTyr 156

DB 237 CGGGCTACTCACTCTTCTGTCGATCACTGACATGACATGACCTTATTAATTAAGTAT 236

QY 157 SerAsnGlnLeuAsnProHisCysValSerLeuTyrIleuSerSerProGluAspAsp 176

DB 297 AGTAACCAAGAATCAACACTGTGTGCTCTTCAAGCTATCAAGTCCCTGAAGATGAC 356

QY 177 ProThrCysValThrIleGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuPro 196

DB 357 CCAACTGCAAAACAAAGAAATTTGGGCCACCACTTTGATTCACAGGTCTCTTCT 416

QY 197 AspTyrThrProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 216
 DB 417 GACTATACCTCCCTCCAGAAATTTCTCTTTGAAAGACACACGATTCGATTCATTTGATGCG 476
 QY 217 MetLeuTyrLeuProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 236
 DB 477 ATGCTCTACAGCCTTATATATCTACAGCCTCGAAGAAATATCTCTACTGCTGTTCTATCA 536
 QY 237 TTYGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 256
 DB 537 TATGGTGGTCTCAGGTGCGATTGGTGAATATCGGTTTAAAGAGATCAAGTATTTCCCG 596
 QY 257 LeuSerThrLeuAlaSerLeuGlyTyrValValValLysLysAsnArgLysSerCys 276
 DB 597 TTGAATACCTTACCTCTAGGTTATGTTGTGTGTATGACACACAGGGGATCTCTGT 656
 QY 277 HisArgGlyLeuLysPheGlnGlyAlaPheLysTyrLysMetGlyGlnIleGluLeuAsp 296
 DB 657 CACCGAGGGCTTAAATTTGAAAGCCCTTTAATATTAATGAGTCAATAAGAAATTTGAC 716
 QY 297 AspGlnValGlnGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 316
 DB 717 GATCAGGTGGAAGGACCTCCAAATATCTAGCTTCTCGATATGATTCATTCATTCAGTACG 776
 QY 317 -ValGlyIleHisGlyTyrSer-Tyr-GlyGlyTyrLeuSer 329
 DB 777 TGTGGGATCCACCGCTGGGCTCTATGGAGATACCTCTCC 818

RESULT 9
 AK050021

LOCUS 3327 bp mRNA linear HTC 05-DEC-2002
 DEFINITION Mus musculus adult male liver tumor cDNA, RIKEN full-length
 enriched library, clone:730003D12 product:DIPEPTIDYL PEPTIDASE 9
 homolog [Homo sapiens], full insert sequence.
 AK050021
 AK050021.1 GI:26340743
 HTG, CAP trapper.
 Mus musculus (house mouse)
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 3
 11042159

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Komno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Hatada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, K., Ozawa, K., Tanaka, T., Matsunaga, S.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Kira, A. and Hayashizaki, Y.
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 4
 20530913
 11076861

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 4 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Arakawa, T., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,

Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiya, H.,
 Kuehl, P., Lewis, S., Matsuo, Y., Nkaido, I., Pesole, G.,
 Quackenbush, J., Schriml, L. M., Staball, F., Suzuki, R., Tomita, M.,
 Wagner, L., Wasth, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bona, M. F., Brownstein, M. J., Bult, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
 Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombaur, P., Nordone, P.,
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
 Toyohara, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
 Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.
 and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 21085660
 11217851

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 5 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420 563-573 (2002)
 6 (bases 1 to 3327)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hirazawa, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,
 Koya, S., Kurimura, C., Matsuyama, T., Miyazaki, A., Murata, K.,
 Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissue was provided by William A. Held, Roswell Park Cancer
 Institute, Department of Molecular and Cellular Biology, Elm and
 Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully
 acknowledge.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/.
 Location/Qualifiers
 1. 3327
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM:DB:C730003D12"
 /db_xref="taxon:10090"
 /clone="C730003D12"
 /sex="male"
 /cell_type="tumor"
 /tissue_type="liver"
 /clone_id="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 176..2764

CDs

/note="unnamed protein product, DIPEPTIDYL PEPTIDASE 9
homolog [Homo sapiens] (SPT|AAL47179, evidence: FASTY,
92%ID, 99.6%length, match=2580)
putative"

/codon_start=1
/protein_id="BAC34034.1"
/db_xref="gi:26340744"

/translation="MCSGVSPVEQVAADMDJTAFCVQKHSMDGLRSIIHGRKSS
GLIVKAPDFQVQKPDSPGSHRLTYLAMPYGRNSLHYSIIPKVRKALLL
SMKMLDHOATPHHGVYSREBELREKRLGVRTSYDHSSEGLFPLPANSIPLH
CRDGKGNFMTSPMKRLEIKTQCSPPDRPICRADPFPFPIINSDIVANIEGEE
RRLTFCHQSGAGVNDNPKSAGVATFVIOEDRFRFCWMCPTASWESGAKTILY
EEVSESEVLIHVSPPALERTDSYRPTKSGNPKIALKLAELQTHQKIVSCE
KEIVQPSSELPKVEYIARAGMTDQKAMFLDRPQORLQVLPALFIPAVESE
AORQAAAVPRNVOPEVYEVTNVMIVNDIHPFPQAGQODFCFLRANEKTFG
CHLYRVATELTKDYDWTETPLSEDEKCPDKKEVALTSGEMVLSHSGKIVNBO
TKLYVFOCTKDTPLRHHLYVYSASGIYVLTITGSHSSMSQSPMPFHSISVS
YPCVHVTKLSGPDPLHQPFRFWSMWEANCPDVPPEIPFHFRAVQVLYGM1
YKPHLOGRHNPVLYVGGPQVQLVNNSEKGIKYELNLTASIGYAVVVIDGSGC
ORGLHFEGLAKNQMGVIEIDQVGLQVAKRYGIDLSRYAIGHMSYGFSLMGLI
HKPOVFKVLAIGAPVVMAYDTGYETRYMDVPENNQOGYAGSVALHVEKLPENR
LIIHGFLENVHFHTNFIYSLIRAKPQLOLYPERHSIRCRSEGEHYETLILH
FLOEHL"

BASE COUNT 702 a 1015 c 919 g 691 t
ORIGIN

Alignment Scores:

Pred. No.: 5.44e-127 Length: 3327
Score: 1204.50 Matches: 234
Percent Similarity: 53.75% Conservative: 53
Best Local Similarity: 43.82% Mismatches: 72
Query Match: 62.31% Gaps: 175
DB: 11 Indels: 2

US-10-070-464-7 (1-360) x AK050021 (1-3327)

```
QY 2 Gluaspalaargserialgiyvalaathrphvealileugingluinphesaprgtyr 21
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 872 GACAAATCCCAATCAGCAGCGGTGGCCACTTTGTCATCCAGAGAGATTGCACGCTTC 931
QY 22 Sergiytrttrpocyeprolysalagluhrthrprosergiygl---lyslileu 40
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 932 ACTGGGTGCTGTGTGTCGCCACGGCTCTTGGGAGAGCTCCGAAGGTCTCAAGACGCTG 991
QY 41 ArgileuYrYgluYgluasnaspGluSerGluValIleIleHisValIleThrserPro 60
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 992 CGCATCCTATATGAGAGATGAGACAGTCTGAAAGGAGTCATTCACTGCCCCC 1051
QY 61 MetleugluhrnarxargalaaSpSerPheargtyrProlysthrGlyThrIleAsnPro 80
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1052 GCCCTGAGAGAGAGAGACGACACTCTACCGCTACCCCAAGGACAGGACAAAGAACCC 1111
QY 81 LysValIThrPheIyMetSerGluIleMetIleAspIleagluYargIlele----- 98
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1112 AAGATTGCCCTGAGCTGTGAGCTCCAGACGGACATCAAGGCAAAATGCTGTCAAC 1171
QY 98 ----- 98
Db 1172 TGGGAGAGAACTGTACAGCATTCAGCTCCCTTTCCCAAGTGAATGACATCCGC 1231
QY 98 ----- 98
Db 1232 CGGGCTGCTGACACAGGAGCGCAAAATATGCTTGAGCCATGTTCTGACCGTCCAG 1291
QY 98 ----- 98
Db 1292 CAAGCGCTTCAAGCTTGTCTCTGCCCCCTGTCTCTTCATCCGCGCGCTTGAGAGTAG 1351
QY 98 ----- 98
Db 1352 GCCCAGCGGACAGGAGCTGCCAGAGCCGTCCCAAGATATGACAGCCCTTTGTATCTAT 1411
QY 98 ----- 98
```

```
Db 1412 GAAGAGTCAACCAATGCTGTGATCAACGTCCACGACATCTTCCACCGCTTCTCAGGCT 1471
QY 98 ----- 98
Db 1472 GAGGCGCAGCAGACTTTTGTTCCTTCCTGTCACCAAGATGACAGACTGGCTTCTGCCAC 1531
QY 98 ----- 98
Db 1532 CTGTACAGAGGTCAACGTGAACCTTAACCAAGACGTATGACTGAGCGGAACCCCTCAGC 1591
QY 98 ----- 98
Db 1592 CCTACAGAAATGAGTTTAAGTGTCCCATCAAGGAGAGTGTGCCCCCTGACCACTGAGCGAG 1651
QY 99 -----ValaspgluValIarghrleuVal 106
Db 1652 TGGAGAGCTTGTGAGGACATGCTCCAAAGATCTGGGTCAACGACACAGAGAGCTGTG 1711
QY 107 TyrPhegluYThrIyAspSerProleuGluHisIleuYrValIleSerYrVal 126
Db 1712 TACTTTCAAGGTACAAAGACACACACCGCTGAGACATCACTCTATGTGTGCTACGACGAG 1771
QY 127 AsnProgiYgluValIThrArgLeuThrAspArgIyYrSerHisSerCysIleSer 146
Db 1772 TCAGCAGCGCAGATCTGTGCGGCTCACACGCTCGCTTCTCCACAGCTGTCTCCATGAC 1831
QY 147 GluHisCysAspPhePheIleSerYrSerAsnGluIlyAsnProHisCysValSer 166
Db 1832 CAGAGCTTCGACATGTTGTGATGATCACTACAGCAGGTGAGAGACGCCACCTGTGTACT 1891
QY 167 LeuYrIlyLeuSerSerProgluAspAspProThrCysIySthrIySgluPheThrPa 186
Db 1892 GTGTACAGCTGAGGCGGCCCGCGATGATGACCAACGACACAGCAACCAACGCTTCGAGGC 1951
QY 187 ThrIleuAspSerIalgiYProleuProAspYrThrProProgluIlePheSerPhe 206
Db 1952 AGCATGATGAGAGCGCACAATGTGCCCCCAACATGATGTGCCCCCTGAGATCTTCCACTTC 2011
QY 207 GluSerThrThrIyPheThrLeuYrGlyMetLeuYrIlyProHisAspIleuInPro 226
Db 2012 CACACCCGTGACAGCTGACCTCTACCGCATGATCTACAGGCCACACACCTCTGACACT 2071
QY 227 GlyIlySthrYrProThrValIleuPheIleYrGlyIyProgluValIleuValAsn 246
Db 2072 GGGAGGAGACACCCCACTGTCTCTTGTCTATGGGGGCCACAGAGTGCATGTGTGAAC 2131
QY 247 AsnArgPheIySgluValIySerYrPheArgLeuAsnThrIleuIleSerIleuGlyYrVal 266
Db 2132 AACTCCTTTAAGCGCATCAAAATACCTGCGGCTAAATACATGGCATCTTGGGCTATGCT 2191
QY 267 ValIyValIleAspAsnArgIySerCysHisArgGlyLeuIlyPheGluGluIlePhe 286
Db 2192 GTGTGTGATCGATGTGCGGAGCTCTGTGACGGGGGCTGTGACCTTGTGAGGGGGCCTG 2251
QY 287 LysYrIlyMetGlyGluIleGluIleAspAspGluValIleGluIlyLeuGluIlyrIleuAla 306
Db 2252 AAAATATCAATGGGCGCAGGTGAGATTGAGACACAGGAGGAGCTTCCAGTACGTGCT 2311
QY 307 SerArgIyAspPheIleAspLeuAspArgValIleGlyIleSgluYrIyPserYrIyGly 326
Db 2312 GAGAAATATGCTTCACTTGAATCTTGAGCGCAATGCGCATTCATGTGTGCTCTCAACGGCGGC 2371
QY 387 TyrIleuSerIleuMetAlaIleuMetGluIleArgSerAspIlePheArgValIleIleIle 346
Db 2372 TTCTCTACATCATGGGCTCATTCACAGACCAAGAGTTCATAGGTCAAGTTCGCGGC 2431
QY 347 AlaProValIThrLeuIlyPheIlePheYrAspThrIyYrThr 360
Db 2432 GCTCCTGTCACTGTGTGATGAGCGCTATGACACAGAGGTACAG 2473
```

RESULT 10
AK050023

Dp	1112	CCCTTGAGGAGGAGGAGGAGGAGCTCTACCGCTACCCAGAGACAGGAGCAAGAACCC	1171
Qy	81	LysValThrPheIysMetSerGluIleMetIleAspAlaGluIleValGlyIle-----	98
Dp	1172	AAGATTGCCCTGAAGCTGGCTGAGCTCCAGACGAGCATCAGGGCAAAATCGTGTCAAGC	1231
Qy	98	-----	98
Dp	1232	TGGCAGAAAGAACTGTGTACAGCCATTCAAGCTCCCTTTTCCCAAAAGTGAAGTACATGACC	1291
Qy	98	-----	98
Dp	1292	CGGGCTGGCTGGACACGCGGAGCGCAAAATATGCTGGCGCATGTTCTTGAGACCGTCCAG	1351
Qy	98	-----	98
Dp	1352	CAAGGCTTCAGCTTGTCCTCGTCCGCTGCTGCTCATCCGGCCGTTGAGAGTGA	1411
Qy	98	-----	98
Dp	1412	GCCCAAGCGGAGGAGGAGCTTGCCAGAGCCGTCCCAAGAAATGTGAGCCCTTTGTGATCTAT	1471
Qy	98	-----	98
Dp	1472	GAAAGATCACCAATGTGTGATCAAGTCCACGACATCTTCCACCCGTTTCTCAGGCT	1531
Qy	98	-----	98
Dp	1532	GAGGCGCAGCAGACCTTTTGTTCCTTCGTCCCAAGATCAAGACTGGCTTGCAC	1591
Qy	98	-----	98
Dp	1592	CTGTACAGGGTCACAGTGGAACTTAAAAACAAGACTATGACTGAGCGAAACCCCTCAGC	1651
Qy	98	-----	98
Dp	1652	CCTACAGAAATGAGTTTAAGTCCCACTCAAGAGAGGGTCCGCTGACAGTGGCGAG	1711
Qy	99	-----ValAspGluValArgArgIleVal-----	106
Dp	1712	TGGAGAGTCTGTGAGAGCATGGCTCCAAATGTTGGGTCAACGAGCAGACGAGCTGGTG	1771
Qy	107	TyrPheGluGlyThrIleAspSerProLeuGluIleHisIleValIleValIleValIleVal	126
Dp	1772	TACTTTCAGGTATCAAAAGACACACCGGTGAACATCACCTCTGTGTGAGTTCAGCTACAG	1831
Qy	127	AsnProGlyGluValThrArgIleThrAspArgGlyTyrSerHisSerCysCysIleSer	146
Dp	1832	TCAGAGGCGAGATGTGCGGCTCACCGCTCGCTTCTCCACAGCTGGCTCCATGAGC	1891
Qy	147	GlnHisCysAspPheIleSerIleTyrSerAsnGlnIleAsnProHisCysValSer	166
Dp	1892	CAGAGCTTCGACATGTTGTGATGACTACAGCAGTGTGAGACAGCCACCTGTGTATCAT	1951
Qy	167	LeuTyrIleLeuSerSerProGluAspAspProThrCysIleThrIleGluPheThrAla	186
Dp	1952	GTGTACAAAGCTGAGGCGGCCCGATGATACCCACATGCACAAGCAACACGCTTCGGGCC	2011
Qy	187	ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe	206
Dp	2012	AGCATGATGGAGGCGACGATTGGCCCCCAAGACTATGTGCCCCCTGAAGATCTTCCATTC	2071
Qy	207	GluSerThrThrGlyPheThrIleuTyrGlyMetLeuTyrIleProHisAspLeuGlnPro	226
Dp	2072	CACACCGGTGACAGCTGCAGCTCTAGGGCATGTACAGGCCACACACCTTGAACTT	2131
Qy	227	GlyIleIysIleTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn	246
Dp	2132	GCGAGGAAGACCCCACTGCTCTCTTGTCTATGGGGGCCCAACAGGTCAAGTGGTGAAC	2191
Qy	247	AsnArgPheIysGlyValIleTyrPheArgIleAsnThrIleAlaSerIleGlyTyrVal	266
Dp	2192	AACCTCTTTAAAGGAGTAAATCTGTGGCTAAATATACATGGCATCTTGTGGCTATGCT	2251

OY		267	Va Val Val Val leAsPanaRgGlyserCyHisArgIylVenuIsPheGlu ValA Phe	286
Db		2252	G G T G G G ATC CAT GGT CG AGG CTCT CTG CA GGG GGCC TG CA CTTC GA GGG GGCC CTG	2311
OY		287	Lys Tyrl SerMetGc Gly Gln l leG Lull leAsPSPGl nAlS lUgl yLeuG In Tyrl LeuAlA	306
Db		2312	AAAATCAAAATGGGCCACAGGTGGACATTGAGGACCAGGTGGAAGGCTTGCACTGACATGCTGCT	2371
OY		307	SerArG TyraSpPhe lleAsPLeuAsPaRValG yl leH seG lyTPSer Tyrg yGly	326
Db		2372	GAGAAGATATGGCTTCATTTGACTTAGCGCGATGCAGTGCCCATTCATGCTGGTCTTAAGCGCGCG	2431
OY		327	Tyrl eUSer leuMetAlaleuMetGlnArSerSerSp lePheArGValAlAlleAlle Gly	346
Db		2432	TTCTCTTCACATCCATGAGGGCTTCATCCACAACGCCAACATGTTCACAGTACGATGACATTCGCGCG	2491
OY		347	AlaProValThr LeuTrp leuPhe TyraSpThnG ly Thr	360
Db		2492	GCTCTCTGTCACATGTCGTGATGGCTATGACACAGGGTACAG	2533
RESULT 11				
LOCUS	AKO78301			
DEFINITION	AKO78301	3457 bp	mRNA	linear HTC 05-DEC-2002
ACCESSION	AKO78301		Mus musculus adult male olfactory brain cDNA, RIKEN full-length	
VERSION	AKO78301		enriched library, clone: e430584g11 produce: DfPEPTIDb PEPTIDASE 9	
KEYWORDS	AKO78301		homolog [Homo sapiens], full insert sequence.	
SOURCE	AKO78301.1 GI:26347124		HTC; CAP trapper.	
ORGANISM	Mus musculus (house mouse)		Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	Carninci, P. and Hayashizaki, Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL MEDLINE	Meth. Enzymol. 303, 19-44 (1999)			
PUBMED	99279253			
REFERENCE	10349636			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
TITLE	Itoh, M., Komou, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.			
JOURNAL MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)			
PUBMED	20499374			
REFERENCE	11042159			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,			
TITLE	Komou, H., Akiyama, J., Nishi, K., Kitunai, T., Tashiro, H., Itoh, M.,			
JOURNAL MEDLINE	Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,			
PUBMED	Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashinagi, K.,			
REFERENCE	Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M.,			
AUTHORS	Yoneeda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,			
TITLE	Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.			
JOURNAL MEDLINE	RIKEN integrated sequence analysis (RISA) system--384-format			
PUBMED	sequencing pipeline with 384 multicapillary sequencer			
REFERENCE	Genome Res. 10 (11), 1757-1771 (2000)			
AUTHORS	4			
TITLE	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,			
JOURNAL MEDLINE	Arkawa, T., Hara, A., Fukunishi, Y., Komou, H., Adachi, J., Fukuda, S.,			
PUBMED	Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,			
REFERENCE	Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,			
AUTHORS	Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,			
TITLE	Flatschmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H.,			
JOURNAL MEDLINE	Kuchel, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,			
PUBMED	Quackenbush, J., Schriml, L.M., Staab, F., Suzuki, R., Tomita, M.,			
REFERENCE	Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,			
AUTHORS	Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,			
TITLE	Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bull, C.,			

TITLE
JOURNAL
MEDLINE
PubMed
REFERENCE
AUTHORS
5
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3457)

JOURNAL
REFERENCE
AUTHORS
6
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shitaki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takekida, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Yumatsuru, M., and Hayashizaki, Y.

TITLE
JOURNAL
COMMENT
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tezumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES
source
1. 3457
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM_DB:6430584G11"
/db_xref="taxon:10090"
/clone="6430584G11"
/sex="male"
/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
182..2770

CDS

/note="unnamed protein product; DIPEPTIDYL PEPTIDASE 9 homolog [Homo sapiens] (SPTR|AAL47179, evidence: PASTY, 9281D, 99.6%length, match=2580)
putative"
/codon_start=1
/protein_id="BAC37211.1"
/db_xref="GI:26347125"
/translation="MCSGSPVPEQVAVAGMDYARFCVQKSHMDGLRSLIHSGRSKSS
GLIVSKAPHDFOVOKPDESGPSHRLYYLYGMPYGRRENLIASEIPKRYKALILL
SKWMLDHPQATPHHGVYSREELRLERKGLVGTISYDPSHSGPLFOANSLIFH
CRDGGKNGRVSPMKPLEIKTQCSGPRMDKICPADPAPFSFINNSDVIANIEEIEE
RLIFCHQSGAGVLDNPKSAGVAIFVIOEFDFRPTGCMWCPITASWESBGLKRLILY
EEVDESEVEIVHVPSPALERKTDSTYRPTGSKNPRIALKEIQTDHQKIVSSCE

BASE COUNT 766 a 1033 c 926 g 732 t
ORIGIN
Alignment Scores:
Pred. No.: 5,77e-127 Length: 3457
Score: 1204.50 Matches: 24
Percent Similarity: 53.75% Conservative: 53
Best Local Similarity: 43.82% Mismatches: 72
Query Match: 62.31% Indels: 175
DB: 11 Gaps: 2
US-10-070-464-7 (1-360) x AK078301 (1-3457)

QY 2 GUASPAAArgSerAlaGlyValAlaThrPheValLeuGlnGluIubPheAspArgTyr 21
+++++
DB 878 GACAAATCCCAATTCAGACGAGCGCTGGCCACTTTGTATCATCAGAGAGAGTTCAGCCGCTTC 937
QY 22 SERGIYTYTTPITPCySProlYsalGluThrThrProSerGlyGly--LysIleLeu 40
+++++
DB 938 ACTGGGTGTCTGGTGGTGGCCCGCGCTCTTGGGAGAGGCTCCGAAAGCTTCAGAGCGCTG 997
QY 41 ArgIleLeuTyrGlnGluIubAspGluSerGluValGluIleLysValThrSerPro 60
+++++
DB 998 CGATCTCTATATAGAGAGAGTGAAGCGCTGAGAGTGAAGTGAATGATATAGTCCCTCCGCC 1057
QY 61 MetLeuGluThrArgArgIlaAspSerPheArgTyrProLysThrGlyThrAlaAspPro 80
+++++
DB 1058 GCCCTGAGAGAGAGAGAGAGAGAGTCTCTACCGCTTACCCAGAGAGAGAGAGAGAGAGAGAG 1117
QY 81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleLys 98
+++++
DB 1118 AAGATTGCCCTGAGAGTGTCTGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1177
QY 98 ----- 98
DB 1178 TGGCAG 1237
QY 98 ----- 98
DB 1238 CGGCGTGGCTGAGACACGGAG 1297
QY 98 ----- 98
DB 1298 CAAAGGCTTCAAGTGTCT 1357
QY 98 ----- 98
DB 1358 GCCCAG 1417
QY 98 ----- 98
DB 1418 GAAGAAGTCAACATGTCTGATCAACGTCACAGACATCTTCCACCGGTTTCTCTCAGGCT 1477
QY 98 ----- 98
DB 1478 GAGGCGCAGACAGAGACTTTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1537
QY 98 ----- 98
DB 1538 CTGTACAGAGGTCACTGAGACTTAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
QY 98 ----- 98


```

Db      1598 CCTACAGAAAGATGAGTTTAAGTCCCATCAAGAGAGGTCGCCCTGACCAAGTGGCGAG 1657
QY      99 -----ValAspGluValArgArgLeuVal
Db      1658 TGGGAGGCTTGTTCAGAGCATGGCTCCAAAGATCTGGGTCAAGAGCAAGCAAGCTGGTG 1717
QY      107 TyrPheGluGlyThrLysAspSerProLeuGluHisLysLeuValValSerTyrVal 126
Db      1718 TACTTTCAAGGTACAAAGAGACACCGCTGGAAATCACTTATGTGTGTCAGTACAG 1777
QY      127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysLysLeu 146
Db      1778 TCAGCAGGCGAGATGTGGGTGCTCAACGCTGGCTTCCCAAGCTGCTCCCAATGAGC 1837
QY      147 GlnHisCysAspPhePheLysSerLysTyrSerAsnGlnLysAsnProHisCysValSer 166
Db      1838 CAGAGCTTCGACATGTTGTGTAGTCACTACAGCACTGACAGACCGCACCCTGTGTACT 1897
QY      167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPala 186
Db      1898 GTGTCAAGCTGAGCGGCCCGCATGATGACCCATGACCAAGCAACCAAGCTTCTGGGCC 1957
QY      187 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe 206
Db      1958 AGCATGATGAGAGCGACGCCAATTGCCGCCAGACTATGTGCCGCCCTGAGATCTTCCACTTC 2017
QY      207 GluSerThrThrGlyPheThrLeuTyrLysMetLeuTyrLysProHisAspLeuGlnPro 226
Db      2018 CACACCCGTCGACAGCTGACGTCTACGSCATGATCTACACACCAACCTGCAACT 2077
QY      227 GlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGluValGlnLeuValAsn 246
Db      2078 GGGAGGAAGACACCCCATGCTGTCTTGTCTATGGGGGCCCAAGGTGACAGTTGGTGAAC 2137
QY      247 AsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266
Db      2138 AACCTCCTTAAGGGATCAAAATACCTGCGCTAAATACCTGCGACCTTGGCTATGTCT 2197
QY      267 ValValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGlyAlaPhe 286
Db      2198 GTGGTGGTATGATGATGCTGGGGCTCTCTCAACCGGGGCTCTCACTTCAAGGGGCCCTCG 2257
QY      287 LysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 306
Db      2258 AAAAATCAATGGCGCAGAGGTGAGATTGAGACCAAGGTGAAAGGCTTGCAGTACCGTCT 2317
QY      307 SerArgTyrAspPheLysLeuAspArgValGlyIleHisGlyThrSerTyrGlyGly 326
Db      2318 GAGAAGTATGGCTTCACTTACCTTACGACCGAGTCGCCATCATGCTGCTTACCGCGGC 2377
QY      327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346
Db      2378 TTCTCTCATCTATGAGGGCTCATCCCAAGCAAGCAAGTTCAGGTATCCATTTGGCGGC 2437
QY      347 AlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
Db      2438 GCTCCGTGCTGCTGTGTGATGGCTATGACACGAGGTACACG 2479

RESULT 12
BU239476 1082 bp mRNA EST 26-NOV-2002
LOCUS BU239476 60322338F1 CSEQCHN33 Gallus gallus cDNA clone CHEST24919 5', mRNA
DEFINITION sequence.
ACCESSION BU239476 GI:25485538
VERSION BU239476.1
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1082)
REFERENCE Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
AUTHORS

```

```

TITLE Pong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
JOURNAL A Comprehensive Collection of Chicken cDNAs
MEDLINE Curr. Biol. 12 (22), 1965-1969 (2002)
PUBMED 22335534
COMMENT 012445342
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
source
1..1082
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHST24919"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN33"
/note="Organ: Liver; Vector: pBluescript II KS(+); Site_1:
Score: 1; Site_2: Not1; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was bluntend, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
BASE COUNT 289 a 240 c 280 g 273 t
ORIGIN
Alignment Scores:
Pred. No.: 0 3.05e-118 Length: 1082
Score: 1122.00 Matches: 217
Percent Similarity: 88.01% Conservative: 18
Best Local Similarity: 81.27% Mismatches: 27
Query Match: 58.04% Indels: 5
DB: 13 Gaps: 2
US-10-070-464-7 (1-360) x BU239476 (1-1082)
QY 97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 116
Db 107 ATCTATGTGATGAAAGCAAAAGACTGTCTATTTTCAAGGACGAAAGACTTCTCTTTA 166
QY 117 GlnHisLysLeuTyrValValSerTyrValAspProGlyGluValThrArgLeuThrAsp 136
Db 167 GAACATCACTTGTAGCTTGTACTACACAGAAATCCCGAAGAGTAAGCAAGCAAGCAAGAA 226
QY 137 ArgGlyTyrSerHisSerCysLysLeuGlnHisCysAspPhePheLysSerTyr 156
Db 227 CGTGGTACTCTCAAGCCCTGCTGTGCACCGAGATGTGACATGTTTATCAACCAAGTAC 286
QY 157 SerAsnGlnLysAspProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 176
Db 287 AGCAATCAAGAGACCAACACTGTGTGTCTTACCGGCTGACAGGACATGAAAGATGAT 346
QY 177 ProThrCysLysThrLysGluPheTyrPalaIleLeuAspSerAlaGlyProLeuPro 196
Db 347 GCAAGTCAAAGACAAAGAAATTCGGGTCACAATTTGGATTCAACAGGCGCTCTTCCCT 406
QY 197 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLysTyrGly 216

```



```

Db      407 GATTACATTCCCCAGAGTGTTCCTTTAGAGAGCTCCAGGGGTTTACTATAGG 466
Qy      217 MetLeuTyrLysProHisAspLeuGlnProGlyLysTyrProThrValLeuPheIle 236
         .....:
Db      467 ATGATGTAACAACCTCACAATCTGCACCTGGAAAGAGTACCCTACTGATCTTCATC 526
Qy      237 TTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 256
         .....:
Db      527 TATGAGAGGCTTCAGGTGAGCTAGTGAACAATCGATTTAAAGAGCTCAAAATATTC 586
Qy      257 LeuAsnThrLeuAlaSerLeuGlyTyrValValValValValValValValValVal 276
         .....:
Db      587 TTGAACACCTTGCGCTTTAGGCTTACCTGTTGCTGTTATGACACCGGGGCTTCGC 646
Qy      277 HisArgGlyLeuLysPheGlnGlyAlaPheLysTyrLysMetGlyGlnIleGlnLeu 296
         .....:
Db      647 CACCGAGGCTGAGTTTGAAGAGGCTTTAAATACAAATAGGACAAATAGAAATTTGAT 706
Qy      297 AspGlnValGlu-GlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAsp 316
         .....:
Db      707 GACCAAGTGAAGGCTGCACTACTTGGCGCTCCAGTATGACTTCATGATTTGGCTTCG 766
Qy      316 GValGlyIleHisGlyTyrSerTyrGlyGlyTyr--LeuSerLeuMetAlaLeuMetG 335
         .....:
Db      767 GGTGGCATTGATGGGTGCTCCATGAGAGGCTAACTCTCTTATGAGGCTTTATATGCC 826
Qy      335 nArgSer AspIlePheArgValAlaIleAlaGlyAla-ProValThrLeuTyr--Ile 353
         .....:
Db      827 GAGGTGAGATATCTTCAGAGGTTGCCATCGCGATCACCGGTAAOCTGGGGAGATT 886
Qy      354 PheTyrAspThrGly 358
         .....:
Db      887 CCTATGACACCGGCT 901

```

RESULT 13
 BE888665 674 bp mRNA linear EST 20-OCT-2000
 LOCUS 601513061F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914582 5',
 DEFINITION mRNA sequence.
 ACCESSION BE888665
 VERSION BE888665.1 GI:10345197
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 674)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapds-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
 http://image.llnl.gov
 Plate: LHAM9736 row: 1 column: 15
 High quality sequence stop: 662.
 Location/Qualifiers
 1..674
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3914582"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH_MGC_71"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.1 Kb."

```

BASE COUNT 185 a 143 c 148 g 198 t
ORIGIN
Alignment Scores:
Pred. No.: 5,52e-118 Length: 674
Score: 1117.00 Matches: 219
Percent Similarity: 96.90% Conservative: 0
Best Local Similarity: 96.90% Mismatches: 6
Query Match: 57.79% Indels: 4
DB: 10 Gaps: 0

```

US-10-070-464-7 (1-360) x BE888665 (1-674)

```

Qy      115 ProLeuGlnHisIleLeuTyrValValSerTyrValAsnProGlyGlyValThrArgLeu 134
         .....:
Db      2 CCTTTAGAGCATCAGCTGACCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 61
Qy      125 ThrAspArgGlyTyrSerHisSerCysIleSerGlnHisCysAspPheIleSer 154
         .....:
Db      62 ACTGACCGTGGCTACTACATCTTCTGTCATGACGACGACGACGACGACGACGACGACG 121
Qy      155 LysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGln 174
         .....:
Db      122 AAGTATAGTAACACAGAGATCCACACTGTGATCCCTTTACAGGCTATCAAGTCTCGAA 181
Qy      175 AspAspProThrCysLysThrLysGlnPheTyrAlaThrIleLeuAspSerAlaGlyPro 194
         .....:
Db      182 GATGACCCAACTGGCAAAACAAAGAAATTTGGGCCACCATTTTGATTCAGAGGTCCT 241
Qy      195 LeuProAspTyrThrProGlnIlePheSerPheGlnSerThrThrGlyPheThrLeu 214
         .....:
Db      242 CTTCCTGACTATACCTCCCAAAATTTCTCTTTGAAAGTACTGATTTACATTG 301
Qy      215 TTTGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysTyrProThrValLeu 234
         .....:
Db      302 TATGGATGCTCTACAGGCTCATGATCTACAGCCTGGAAAGAAATATCTCTAGTGTG 361
Qy      235 PheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyr 254
         .....:
Db      362 TTCATATATGATGTGCTCTCAGGTGCAAGTGTGAATATCGGTTTAAAGAGCAAGTAT 421
Qy      255 PheArgLeuAsnThrLeuAlaSerLeuGlyTyrValValValValValValValValVal 274
         .....:
Db      422 TTCGGTGAATACCTTACCTCTCTAGGTTATGTGTTGTGTATGATAGCAACAGGGGA 481
Qy      275 SerCysHisArgGlyLeuLysPheGlnGlyAlaPheLysTyrLysMetGlyGlnIleGlu 294
         .....:
Db      482 TCCTGTACCGAGGCTTAAATTTGAAGGCGCTTTAAATATTAATGAGTCAACATAGAA 541
Qy      295 IleAspAspGlnValGlnGly-LeuGlnTyrLeuAlaSerArgTyrAspPheIleAsp 314
         .....:
Db      542 ATTAGACATCAGGTGAGAGAACTCCAAATATCTAGCTTTCGATATGATTTCTTACTT 601
Qy      314 uAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeu 334
         .....:
Db      602 AGATGTGTGGGATCCAGCTGAGTGGTCTATGAGAGATCTC-TCCCTGATGCAATTAAT 659
         .....:
Qy      334 cGlnArgSerAspIle 339
         .....:
Db      660 GCAAG-TCAGATATC 674

```

RESULT 14
 BM910838 1090 bp mRNA linear EST 12-MAR-2002
 LOCUS BM910838
 DEFINITION AGENCOURT 6615934 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454402
 5', mRNA sequence.
 ACCESSION BM910838
 VERSION BM910838.1 GI:19361217
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1090)
AUTHORS NIH-MGC <http://mgi.mgi.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: L10M1949 row: h column: 19
High quality sequence stop: 622.
Location/Qualifiers

FEATURES

source

```
1..1090
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5454402"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_98"
/Note="Organ: brain; Vector: pOTB7, Site_1: XhoI, Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
1 others
```

BASE COUNT 317 a 254 c 240 g 278 t 1 others

ALIGNMENT SCORES:

Align. No.: 1,11e-116 Length: 1090
Score: 1108.50 Matches: 237
Percent Similarity: 77.92% Conservative: 10
Best Local Similarity: 74.76% Mismatches: 54
Query Match: 57.35% Gaps: 16
DB: 12 Gaps: 4

US-10-070-464-7 (1-360) x BM910838 (1-1090)

```
QY 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluGluPheAspArg 20
DB 82 GAAGAGATGCGAGTACGAGTCCGCTTGTCTCCAGAGAAATTGATGGA 141
QY 21 TysSerGlyIYrTPTTCySProlYsAlaGluThrProSerGlyGlyLysIleLeu 40
DB 142 TATTCGTGCTATGCTGTGCTCCAAAGCTGAAACACTCCACAGTGTGTAATAATCTT 201
QY 41 ArgIleLeuYrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
DB 202 AGAATTCTATATGAAGAAATGATGATCTGAGGTGGAAATTATTCAGTTACATCCCT 261
QY 61 MetLeuGluThrArgAlaAspSerPheArgYrProLysThrGlyThrAlaAsnPro 80
DB 262 ATGTTGGAACAAGAGGAGGAGATTCATTCGTTATCTTAACACGAGATTTCAGAGT 321
QY 81 LysValThrPheLysMetSer-----GluIleMetIleAspAlaGlu 94
DB 322 CCTATCAAGAGAGATGACATTAACAGTGTGATGAGTGTCTTGGCCGCGCATGGA 381
QY 95 GlyArgIleIleValAspGluValArgArgLeuValYrPheGluGlyThrLysAspSer 114
DB 382 TCTATATCAAGTGTGAGAGTCCAGAGGCTGGATATTTTGAAGGACCAAAACATCC 441
QY 115 ProLeuGluHisLysLeuYrValValSerYrValAsnProGlyGluValThrArgLeu 134
DB 442 CCTTAGAGCATCACTGATAGTCACTTAAGTATCCGAGAGAGTGAACAAGGCTG 501
```

```
QY 135 ThrAspArgGlyYrSerHisSerCysGlyIleSerGlnHisCysAspPheIleSer 154
DB 502 ACTGACCGTGGCTACTCATCTTGTGCTGACATCACTGACACTGACTTCTTATAGT 561
QY 155 LysYrSerArgGlnLysAsnProHisCysValSerLeuYrLysLeuSerSer-ProG1 174
DB 562 AAGTATATTAACCAAAAGAAATTCACACGTGTGCTTTAAAGCTATCAAGTCCCTGA 621
QY 174 uAspAspProThrCysLysThrLysGluPheThrAlaThrIleLeuAspSerAlaGly-P 194
DB 622 AGATACCCCACTTCCAAACAAAGAAATTTGGGCCACCATTTTGATTCACGACAGNCC 681
QY 194 rLeuProAspYrThrProProGluIle-PheSerPheGluSerThrThrGly-PheTh 213
DB 682 CTCTTCTGACTATCTCTCCACAGAAATTTTCTTTTGAAGATCACTACCGAAATTAC 741
QY 213 rLeuYrGlyMetLeu-YrLysProHisAspLeuGluProGlyLysLysYrProThrY 233
DB 742 ATTGATAGGAGTCTTACCAAGCTCAAGATTAAGCTGAGAAAGAAATATCTTACTG 801
QY 233 allPhe-1IeYrGlyYrProGluValGlnLeuVal-AsnAsnArgPheLysGlyVa 252
DB 802 TGCTTGTCCATATAGGGGGGCTCTCCGGGGCAGTTGGCGGAAAAACCGCTTAAAGAT 861
QY 252 lLysYrPheArgLeuAsnThrLeuAlaSerLeu---GlyYrValValValIleAs 271
DB 862 CAAGTATTTCCCTTGAAGAAACCTCAACCCCTTAAAGTATTGGGGTGGCAGCATTAAC 921
QY 271 pAsnArgGly---SerCysHisArgGly---LeuLysPheGluGluValaPheLysYrLys 289
DB 922 ACAGAGCGCGGTCTGCCACCCAGCCGTTTAACTTATAGAGGGCCCTTGAATATAA 981
QY 289 smetGlyGlnIleGluIleAspAspGluValaGluGly 301
DB 982 CAAAGCCGCCAAGACACACAGACTACTCGCGGGT 1018
```

RESULT 15

BG165205 639 bp mRNA linear EST 06-FEB-2001
LOCUS 602344074P1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:445387 5',
DEFINITION mRNA sequence.

ACCESSION BG165205.1 GI:12671908

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: L10M10245 row: d column: 20
High quality sequence stop: 639.
Location/Qualifiers

FEATURES

source

```
1..639
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:445387"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_89"
/Note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI;
```

Site 2: Salt: Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."

BASE COUNT 169 a 137 c 142 g 191 t
ORIGIN

Alignment Scores:

Pred. No.:	6.03e-116	Length:	639
Score:	1099.00	Matches:	210
Percent Similarity:	98.59%	Conservative:	0
Best Local Similarity:	98.59%	Mismatches:	2
Query Match:	56.85%	Indels:	2
DB:	10	Gaps:	0

US-10-070-464-7 (1-360) x BG165205 (1-639)

```
QY      140 SerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAenGln 159
      |||
Db      2  TCACATTCCTTGCTGCATCAGTCAGCAGCTGACTCTTTATAGTATAGTATACAG 61
QY      160 LysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCys 179
      |||
Db      62 AAGAATCCACACTGTGTGTCCTTTACAAGCTATCAAGTCTGAGATGACCACTTGC 121
QY      180 LysThrLysGluPheThrPheThrIleLeuAspSerAlaGlyProLeuProAspTyrThr 199
      |||
Db      122 AAAACAAGAAATTTGGGCAACCAT-TTGATTCAGCAGGTCCTTCTGCACTACT 180
QY      200 ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 219
      |||
Db      181 CCTCCAGAAATTTTCTTTGAAGTACTACTGATTTACATTTGATGGATGCTCTAC 240
QY      220 LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly 239
      |||
Db      241 AAGCCCTCATGATCTACAGCCTGGAAGAAATCTCTACTGTGCTGTCATATATGTGTGT 300
QY      240 ProGlnValAlaLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAenThr 259
      |||
Db      301 CTTCAAGTGCAGTGTGTGAATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACC 360
QY      260 LeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCysHisArgGly 279
      |||
Db      361 CTAGCCTCTCTAGGTTATGTGTGTAGATAGACACAGGGGATCTGTCACCGAGGG 420
QY      280 LeuLysPheGlyGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnVal 299
      |||
Db      421 CTTAAATTTGAAGGCCCTTTAAATATTAATGGGTCAAATAGAAATTAACGATCAGGTG 480
QY      300 GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
      |||
Db      481 GAAGGACTCCATATCTAGCTTCTCGATATGATTTCACTGACTTATGATCGTGTGGGCAATC 540
QY      320 HisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle 339
      |||
Db      542 CACGGCTGTGCTCATGAGGATACCTCTCCGATGCAATTAATGACAGAGTCAGATATTC 600
QY      340 Phe-ArgValAlaIleAlaGlyAlaProValThrLeu 351
      |||
Db      601 TTCAGCGTTGCTATGTGAGGCCCACTACTCTG 637
```

Search completed: October 16, 2003, 03:25:54
Job time : 2181.86 secs


```

PR 10-SEP-1999; 99AU-0002762.
PR 18-FEB-2000; 2000AU-0005709.
XX
PA (UNSY ) UNIV SYDNEY.
XX
PI Abbott CA, Gorell MD;
XX
DR WPI; 2001-281520/29.
DR N-PSDB; AAC85694.
XX
PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT substrates, identifying inhibitors of DPP8 catalytic activity which
PT have therapeutic uses, and for detecting activated T cells
XX
PS Claim 1; Fig 2; 78pp; English.
XX
CC This sequence represents human dipeptidyl aminopeptidase (DPP8).
CC DPP8 has substrate specificity for H-Ala-Pro-pNA, H-Gly-Pro-pNA and
CC H-Arg-Pro-pNA. Therefore, it is a prolyl oligopeptidase and a
CC dipeptidyl peptidase, because it is capable of hydrolysing the
CC peptide bond C-terminal to Pro in each of these compounds. DPP8
CC is homologous with human DPPIV. DPP8 is useful for cleaving a
CC substrate, and for detecting an activated T cell which involves
CC measuring the level of DPP8 gene expression in a T cell. The level
CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
CC in the cell. It is also useful for identifying a molecule capable
CC of inhibiting the cleavage of the substrate by DPP8. Molecules
CC identified as inhibiting DPP8 catalytic activity may be useful for
CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC in non-insulin dependent diabetes mellitus and other disorders
CC involving glucose intolerance, enhancing mucosal regeneration and
CC as immunosuppressants.
XX
SQ Sequence 882 AA;
Query Match 100.0%; Score 4700; DB 22; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAMETEQVEIFETADCEENIESQDRPKLEPPYVERYSMSQKLLADTRKHYGM 60
Db 1 MAAMETEQVEIFETADCEENIESQDRPKLEPPYVERYSMSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVFRNDPFGHSRIRIYLLAMSGENRENTLYSIPKTTNRAAVLMSKPL 120
Db 61 AKAPHEMFVFRNDPFGHSRIRIYLLAMSGENRENTLYSIPKTTNRAAVLMSKPL 120
QY 121 DLFOATLDYGMVSREELLRERKRGIVGASVDYHQSGLFPLPAGSGIYHVDGPG 180
Db 121 DLFOATLDYGMVSREELLRERKRGIVGASVDYHQSGLFPLPAGSGIYHVDGPG 180
QY 181 FTQOPLRBNLVETSCPNIRMDPKLCPADPWIAFIHSDIMISNIVTREERRLTYVNE 240
Db 181 FTQOPLRBNLVETSCPNIRMDPKLCPADPWIAFIHSDIMISNIVTREERRLTYVNE 240
QY 241 ANMEDDASAGVATFVLOEEDRYSGYWVCRAETTPGSGKILRLIYEENDBSEVEI 300
Db 241 ANMEDDASAGVATFVLOEEDRYSGYWVCRAETTPGSGKILRLIYEENDBSEVEI 300
QY 301 TSPMLERRADSPFRPKGTANPKYTFPMSEIMIDABRIIDVIDKELIOPFEIFEGVE 360
Db 301 TSPMLERRADSPFRPKGTANPKYTFPMSEIMIDABRIIDVIDKELIOPFEIFEGVE 360
QY 361 YIARAGWTPGKYAMSILLDRSQTRLQIVLISPFLFIVEDDVMERQLISVDPSTPL 420
Db 361 YIARAGWTPGKYAMSILLDRSQTRLQIVLISPFLFIVEDDVMERQLISVDPSTPL 420
QY 421 IIVETTDIWINIHIDIEHVPSQSHRETEFIASCKCKGFHLKYTISILSKSKRSSG 480
Db 421 IIVETTDIWINIHIDIEHVPSQSHRETEFIASCKCKGFHLKYTISILSKSKRSSG 480
QY 481 GLPAPDFKCPIKEIEAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTQDSPLEHLLYVS 540
Db 481 GLPAPDFKCPIKEIEAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTQDSPLEHLLYVS 540

```

```

Db 481 GLPAPDFKCPIKEIEAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTQDSPLEHLLYVS 540
QY 541 YVNPGEVTRILDRGYSHSCCISQCHDFFISKYSNOKAPHCVSLYKLSPPEDDPCKTKEF 600
Db 541 YVNPGEVTRILDRGYSHSCCISQCHDFFISKYSNOKAPHCVSLYKLSPPEDDPCKTKEF 600
QY 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIQPKKYPVLIYGGPOVL 660
Db 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDIQPKKYPVLIYGGPOVL 660
QY 661 VNNRFKGVKFRNLNTLASLGIVVVVINDRSGCHRGKKEGAFKYMGIETDDQVEGQY 720
Db 661 VNNRFKGVKFRNLNTLASLGIVVVVINDRSGCHRGKKEGAFKYMGIETDDQVEGQY 720
QY 721 IASRYDFLIDRVRGIGHMSYGYSLMALMQRSDIFRVAIAGAPVTLMIFPDGTTERYM 780
Db 721 IASRYDFLIDRVRGIGHMSYGYSLMALMQRSDIFRVAIAGAPVTLMIFPDGTTERYM 780
QY 781 GHPDQNEQGYLGSVAMQAEKFPSEPNRLILHGFLEENVFAHTSILSLFVRAKPYD 840
Db 781 GHPDQNEQGYLGSVAMQAEKFPSEPNRLILHGFLEENVFAHTSILSLFVRAKPYD 840
QY 841 LQIYPOERHSIRVPSGHEVHLHLHYLOEWLGRIALAKVI 882
Db 841 LQIYPOERHSIRVPSGHEVHLHLHYLOEWLGRIALAKVI 882
RESULT 2
AAE24170
ID AAE24170 standard; Protein: 882 AA.
AC AAE24170;
DT 23-SEP-2002 (first entry)
DE Human dipeptidyl peptidase 8 (DPP8) protein.
KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; Human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme.
OS Homo sapiens.
PN WC0200234900-A1.
PX 02-MAY-2002.
PD 29-OCT-2001; 2001WO-AU01388.
PF 27-OCT-2000; 2000AU-0001078.
PR (UNSY ) UNIV SYDNEY.
PA (UNSY ) UNIV SYDNEY.
PI Abbott CA, Gorell MD;
XX
DR WPI; 2002-454646/48.
DR N-PSDB; AAD38956.
XX
PT New dipeptidyl peptidase (DPP) peptidase, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection.
XX
PS Example; Fig 1; 91pp; English.
CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP proteins are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP8 protein.

```


QY	24	ANNEEDRRSGVATFVLQSEFDPDYSYMMCPAEFTTPSGAKLTILYENDESEVEIIHV	300
Db	241	ANNEEDRRSGVATFVLQSEFDPDYSYMMCPAEFTTPSGAKLTILYENDESEVEIIHV	300
QY	301	TSPMLETRRADSRFYKPTGTANPKVYFKMSEIMDABGRIIDVIDKELIQEFELLFEGVE	360
Db	301	TSPMLETRRADSRFYKPTGTANPKVYFKMSEIMDABGRIIDVIDKELIQEFELLFEGVE	360
QY	361	YIARAGMTPEGKXAMSTILDRGSTRQIVLISPELFIPEDDVMERQGLISSVDSVTPPL	420
Db	361	YIARAGMTPEGKXAMSTILDRGSTRQIVLISPELFIPEDDVMERQGLISSVDSVTPPL	420
QY	421	IYYEETDIIINIHIDIFHVFPQSHHEEIEFIFASECKTGFRHLKYITSIILKESYKXSSG	480
Db	421	IYYEETDIIINIHIDIFHVFPQSHHEEIEFIFASECKTGFRHLKYITSIILKESYKXSSG	480
QY	481	GLPAPSDPKCPIKEEIAITSGEMEVIGRHGNSIQVDEVARLLVYFEGTDSPLBHLVYVS	540
Db	481	GLPAPSDPKCPIKEEIAITSGEMEVIGRHGNSIQVDEVARLLVYFEGTDSPLBHLVYVS	540
QY	541	YVNPGEVTRLIDRGYSHSCCISQHCDFIISKISNQKNPCHVSLYLSSPEDDPCKTKEF	600
Db	541	YVNPGEVTRLIDRGYSHSCCISQHCDFIISKISNQKNPCHVSLYLSSPEDDPCKTKEF	600
QY	601	WATILDSAGBLPYTPPEPIFSEFSTTGFTLYGMLYKPHDLOFGKKYPTVLFITYGAPQVOL	660
Db	601	WATILDSAGBLPYTPPEPIFSEFSTTGFTLYGMLYKPHDLOFGKKYPTVLFITYGAPQVOL	660
QY	661	VNNRFKVKCFKFRNLNTLASIGYVVVVVDNNGSCHRGKJKEGAFKYMGGIIEIDDVBEIGY	720
Db	661	VNNRFKVKCFKFRNLNTLASIGYVVVVVDNNGSCHRGKJKEGAFKYMGGIIEIDDVBEIGY	720
QY	721	LASRYFDIDLDRGIGWMSYGYLSIMALMORSIDIFRVALAGAPVTLMIFFDTGYTERYM	780
Db	721	LASRYFDIDLDRGIGWMSYGYLSIMALMORSIDIFRVALAGAPVTLMIFFDTGYTERYM	780
QY	781	GHPDNEQGYIYLGSAVMAQAKPPSEPNRLLLHGLDENVHPARTSIILSTLVAKGRPYD	840
Db	781	GHPDNEQGYIYLGSAVMAQAKPPSEPNRLLLHGLDENVHPARTSIILSTLVAKGRPYD	840
QY	841	LOIYPOERHSIRVPESEGHYELLHLYLOENSGRIALAKVI 882	
Db	841	LOIYPOERHSIRVPESEGHYELLHLYLOENSGRIALAKVI 882	
RESULT 4			
AAU74749	ID	AAU74749 standard; Protein; 882 AA.	
AC	AAU74749;		
DT	09-APR-2002 (first entry)		
DE	Human protease PR7S-9 protein sequence.		
KW	Human; protease; PR7S; gastrointestinal; Crohn's disease; cancer;		
KW	cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;		
KW	inflammatory disorder; acquired immunodeficiency syndrome; AIDS;		
KW	cell proliferative disorder; developmental disorder; epilepsy;		
KW	Duchenne muscular dystrophy; epithelial disorder; neurological disorder;		
KW	reproductive disorder; endometriosis.		
XX	Homo sapiens.		
XX	OS		
XX	FN	WO200198468-A2.	
XX	PD	27-DEC-2001.	
XX	PF	13-JUN-2001; 2001MO-US19178.	
XX	PR	16-JUN-2000; 2000US-212336P.	
XX	PR	22-JUN-2000; 2000US-213955P.	
XX	PR	29-JUN-2000; 2000US-215396P.	

Query Match	100.0%; Score 4700; DB 23; Length 882;	Best Local Similarity 100.0%; Pred. No. 0;	Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX	07-JUL-2000; 2000US-216821P.		
PR	14-JUL-2000; 2000US-218946P.		
XX	(INCY-) INCYTE GENOMICS INC.		
PA	Yue H, ⁶ Gilliot ⁶ VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;		
PI	Delegene AM, Baughn MR, Nguyen DB, Lee EA, Hafalila A, Khan FA,		
PI	Walia NK, Yeo MG, Lu DAM, Patterson C, Tang YT, Walsh RT;		
PI	Azimzal Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;		
PI	Kallik DA;		
XX			
PS	Claim 1; Page 140-142; 177pp; English.		
XX			
CC	The present invention relates to twenty one new human proteases,		
CC	referred to as PRTS-1 to PRTS-21. The PRTS polymucleotides and		
CC	polypeptides of the invention are useful in the diagnosis, treatment and		
CC	prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and		
CC	Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and		
CC	myocardial infarction, autoimmune/inflammatory e.g. acquired		
CC	immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell		
CC	proliferative e.g. cancer, developmental e.g. Duchenne and Becker		
CC	muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.		
CC	epilepsy and Alzheimer's disease and reproductive e.g. infertility and		
CC	endometriosis disorders. Numerous other examples of each disorder are		
CC	given in the specification. The present protein sequence represents		
CC	the human protease PRTS-9 protein of the invention.		
XX			
SQ	Sequence 882 AA;		
Query Match	100.0%; Score 4700; DB 23; Length 882;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches	882; Conservative 0; Mismatches 0; Indels 0; Gaps 0		
QY	1 MAAMETBQLGVHIFETADCEENIESODRPLLEPPYERYSQKLADTRKHYGM 60		
DB	2 MAAMETBQLGVHIFETADCEENIESODRPLLEPPYERYSQKLADTRKHYGM 60		
QY	61 AKAPHDNFVVRNDPDGHSRIYYLMSGNRENTLFYSIPTKTINAAVLMISWKRL 120		
DB	61 AKAPHDNFVVRNDPDGHSRIYYLMSGNRENTLFYSIPTKTINAAVLMISWKRL 120		
QY	121 DLFOATLDYGYRSREBELRERKRTIGTVIASYVYHQSQTFLPQASGIVHVDGPGQG 180		
DB	121 DLFOATLDYGYRSREBELRERKRTIGTVIASYVYHQSQTFLPQASGIVHVDGPGQG 180		
QY	181 FTQOPLRPNLVETGCPNIRMDPKICPADPDMIAFIHNSNDIWSINVTREERLLTYVHNL 240		
DB	181 FTQOPLRPNLVETGCPNIRMDPKICPADPDMIAFIHNSNDIWSINVTREERLLTYVHNL 240		
QY	241 ANMEDARSAGVAFVVOEEDRYSGYWMCKAETTPSGKILRLVYENDESEVEIITHV 300		
DB	241 ANMEDARSAGVAFVVOEEDRYSGYWMCKAETTPSGKILRLVYENDESEVEIITHV 300		
QY	301 TSPMLFTRRADSFYRPTGTANPKVTPKMSIMIDAEGRITDIVIDKLIQPEILFEGVE 360		
DB	301 TSPMLFTRRADSFYRPTGTANPKVTPKMSIMIDAEGRITDIVIDKLIQPEILFEGVE 360		
QY	361 YIARAGTTPBEKYAMSLILDSQRRLQVILSPELFIPEVDVWEROGLIESVDSVPL 420		
DB	361 YIARAGTTPBEKYAMSLILDSQRRLQVILSPELFIPEVDVWEROGLIESVDSVPL 420		
QY	421 IIVETTDIWINIHDIHFVFPQSHSEBIEFIASBECKTGFRHLKYTISLSEKXYKSSG 480		
DB	421 IIVETTDIWINIHDIHFVFPQSHSEBIEFIASBECKTGFRHLKYTISLSEKXYKSSG 480		

```

QY 481 GLPAPSDKCPKEEIIATSGEMVLRGHSNIQVDEVRLVYFEGTKDSPLEHLLVYVS 540
DB 491 GLPAPSDKCPKEEIIATSGEMVLRGHSNIQVDEVRLVYFEGTKDSPLEHLLVYVS 540
QY 541 YVNPGEVRLTDRGYSHSCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTKEF 600
DB 541 YVNPGEVRLTDRGYSHSCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTKEF 600
QY 601 WATILDSAGPLDPDYPPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPOVQL 660
DB 601 WATILDSAGPLDPDYPPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPOVQL 660
QY 661 VNNRKGKGYKFFRLNTLASIGYVVVYIDNRGSGHRLKKEFGAKRYMGOIEIDDQVEGLQY 720
DB 661 VNNRKGKGYKFFRLNTLASIGYVVVYIDNRGSGHRLKKEFGAKRYMGOIEIDDQVEGLQY 720
QY 721 LASRDFIDLDVJGHSNYSYGYSLSMALMQRSDIFRVAIAGAPVTLMIFYDTGTERYM 780
DB 721 LASRDFIDLDVJGHSNYSYGYSLSMALMQRSDIFRVAIAGAPVTLMIFYDTGTERYM 780
QY 781 GHPDNEQGYLGSVAMQAEKFPSPBNRLLLHGFIDENVHFAHTSILLSPVRAKGPYD 840
DB 781 GHPDNEQGYLGSVAMQAEKFPSPBNRLLLHGFIDENVHFAHTSILLSPVRAKGPYD 840
QY 841 LOIYPOERHSIRVPESGEHYEHLHYLOENIGSRIAAKVI 882
DB 841 LOIYPOERHSIRVPESGEHYEHLHYLOENIGSRIAAKVI 882

RESULT 5
ID AAG78415
AC AAG78415 standard; Protein; 882 AA.
DT 12-APR-2002 (first entry)
DE Amino acid sequence of 21953 human prolyl oligopeptidase.
KW 21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KW diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW Grave's disease; neuronal disorder; demyelinating disease.
OS Homo sapiens.
PN WO200179473-A2.
PD 25-OCT-2001.
PF 11-APR-2001; 2001WO-US40483.
PR 18-APR-2000; 2000US-197508P.
PA (MILL-) MILLENNIUM PHARM INC.
PI Meyers RA, Williamson M;
DR MPI; 2002-034353/04.
DR N-PSDB; AAH99934.
XX New polypeptides 21953, member of human prolyl oligopeptidase family,
XX useful as diagnostic targets and therapeutic agents for controlling
XX cancer, lymphoma and leukemia -
XX Claim 1; Page 102-103; 12pp; English.
CC This invention relates to an isolated 21953 human prolyl
CC oligopeptidase, which is cytosolic, antidiabetic, antiarthritic,
CC neuroprotective, antihypertoid, dermatological, antipsoriatic,
CC antiasthmatic, ophthalmological, antiinflammatory, nootropic,

```

```

CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
CC antianginal, cardiant, antiatherosclerotic, anorectic and
CC metabolic in its action. Uses include gene therapy, expression or
CC activity of 21953 protein modulator, it is useful for identifying a
CC compound which binds to it and can be used in preventing, treating
CC or detecting a cellular proliferative or differentiative disorder.
CC The 21953 molecules can act as novel diagnostic targets and therapeutic
CC agents for controlling disorders associated with the aberrant activity
CC or degradation of peptide hormones e.g., disorders associated with cell
CC differentiation and proliferation such as cancer, immune function,
CC reproductive, neurological and cardiovascular function. The 21953
CC molecules are thus useful for treating and preventing cellular
CC proliferative and differentiative disorders, haematopoietic neoplastic
CC disorders, immune disorders such as autoimmune diseases, diabetes
CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC neuronal disorders, demyelinating diseases, vascular disorders and
CC metabolism or pain disorders. This sequence represents the amino
CC acid sequence of 21953 human prolyl oligopeptidase.
XX
SQ Sequence 882 AA;
Query Match 100.0%; Score 4700; DB 23; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAAMETQLGVEIFETADCEENIESODRPKLPEYVERYSOLKLLADTRKYGVM 60
DB 1 MAAAMETQLGVEIFETADCEENIESODRPKLPEYVERYSOLKLLADTRKYGVM 60
QY 61 AKAPDFMVKRNDSDGHSRIYYLAMSGERENTLPSSELPKTNRAAVMLSKPRL 120
DB 61 AKAPDFMVKRNDSDGHSRIYYLAMSGERENTLPSSELPKTNRAAVMLSKPRL 120
QY 121 DLFOATLDYGMTSREELLREKRIGTVGIASVDYHQSQTFLFOAGSGIYHVKDGPOG 180
DB 121 DLFOATLDYGMTSREELLREKRIGTVGIASVDYHQSQTFLFOAGSGIYHVKDGPOG 180
QY 181 FTQGPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDNIWISNIVTREERLLTYVHNEI 240
DB 181 FTQGPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDNIWISNIVTREERLLTYVHNEI 240
QY 241 ANMEDDASAGYATVLOEFPRIYGYMCPAETTPSGKTLRIIYENDSEVEIIVH 300
DB 241 ANMEDDASAGYATVLOEFPRIYGYMCPAETTPSGKTLRIIYENDSEVEIIVH 300
QY 301 TSPMLETRRADSFYRPKGTANPKYTFPKMSEIMIDAEGRIDVIDKELIQPEILLFEGVE 360
DB 301 TSPMLETRRADSFYRPKGTANPKYTFPKMSEIMIDAEGRIDVIDKELIQPEILLFEGVE 360
QY 361 YIARAGWTPBCKYAMSILLDRSQTRLQIVLISPPELLPYEDDVMERORLIESVPDSVPL 420
DB 361 YIARAGWTPBCKYAMSILLDRSQTRLQIVLISPPELLPYEDDVMERORLIESVPDSVPL 420
QY 421 IYIETTDIWINIHIFHVFPQSHEEIEFTFASCKTGFRLLYKITSILKSKYRRSG 480
DB 421 IYIETTDIWINIHIFHVFPQSHEEIEFTFASCKTGFRLLYKITSILKSKYRRSG 480
QY 481 GLPAPSDKCPKEEIIATSGEMVLRGHSNIQVDEVRLVYFEGTKDSPLEHLLVYVS 540
DB 481 GLPAPSDKCPKEEIIATSGEMVLRGHSNIQVDEVRLVYFEGTKDSPLEHLLVYVS 540
QY 541 YVNPGEVRLTDRGYSHSCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTKEF 600
DB 541 YVNPGEVRLTDRGYSHSCCISQHCDFPISKYSNQNPKVCVSLYKLSPEDDPTCKTKEF 600
QY 601 WATILDSAGPLDPDYPPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPOVQL 660
DB 601 WATILDSAGPLDPDYPPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPOVQL 660
QY 661 VNNRKGKGYKFFRLNTLASIGYVVVYIDNRGSGHRLKKEFGAKRYMGOIEIDDQVEGLQY 720
DB 661 VNNRKGKGYKFFRLNTLASIGYVVVYIDNRGSGHRLKKEFGAKRYMGOIEIDDQVEGLQY 720

```


QY 721 LASRYDFIDLRVGHGWSYGYLSLMAIMORSDFIRVALAGAPVTLMIFFDTGTERYM 780
 DB 721 LASRYDFIDLRVGHGWSYGYLSLMAIMORSDFIRVALAGAPVTLMIFFDTGTERYM 780
 QY 781 GHDDNQGCGYLLGSVMQAQKPESEPNRLLLHGFIDENHFAHTSILLSPVRAKGPYD 840
 DB 781 GHDDNQGCGYLLGSVMQAQKPESEPNRLLLHGFIDENHFAHTSILLSPVRAKGPYD 840
 QY 841 LGIYPOERHSIRPESGEHEHLLHMYLOENLGSRIAALKVI 882
 DB 841 LGIYPOERHSIRPESGEHEHLLHMYLOENLGSRIAALKVI 882
 RESULT 6
 ABU07720
 ID ABU07720 standard; Protein; 882 AA.
 XX
 AC ABU07720;
 XX
 DT 19-MAY-2003 (first entry)
 XX
 DE Human serine protease HIPHM46.
 XX
 KW Human; enzyme; HIPHM46; serine protease; gene therapy; osteoarthritis;
 KW serine protease activity modulation; dipeptidyl peptidase activity;
 KW musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
 KW Alzheimer's disease; paraspinal nuclear palsy; Huntington's disease;
 KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
 KW irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
 KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
 KW colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
 KW multiple sclerosis.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Region 259..260
 FT /note= "Paired glutamates of the beta propeller domain"
 FT Active-site 739
 FT /label= Catalytic_serine_residue
 FT Active-site 817
 FT /label= Catalytic_aspartate_residue
 FT Active-site 849
 FT /label= Catalytic_histidine_residue
 XX
 GN GB2374869-A.
 XX
 PN 22-JAN-2002; 2002GB-0001404.
 PD 30-OCT-2002.
 XX
 PF 22-JAN-2002; 2002GB-0001404.
 XX
 PR 23-JAN-2001; 2001GB-0001760.
 XX
 PA (GLAXO) GLAXO GROUP LTD.
 XX
 PI Edbrooke MR, Lewis AP;
 XX
 DR WPI; 2003-150703/15.
 DR N-PSDB; ABX12255.
 XX
 PT Identifying modulators of serine protease activity useful for treating
 PT musculoskeletal diseases, by contacting cell expressing a novel serine
 PT protease polypeptide with a compound and monitoring serine protease
 PT activity -
 XX
 PS Claim 10; Page 26-29; 38pp; English.
 XX
 CC The invention relates to a method of identifying a substance that
 CC modulates serine protease activity, comprising contacting a cell such as
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a
 CC virus, expressing a serine protease polypeptide (HIPHM 46), or its
 CC variant having dipeptidyl peptidase activity, or a serine protease
 CC isolated from the cell with a test substance and monitoring for serine

CC protease activity. The method is useful for identifying a substance that
 CC modulates serine protease activity. A modulator of the serine protease is
 CC useful in the manufacture of a medicament for treatment or prophylaxis of
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
 CC infection, Alzheimer's disease, paraspinal nuclear palsy, myotonic
 CC dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
 CC Additional disease that may be treated using modulators of the serine
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung
 CC disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
 CC sclerosis. The present sequence represents the amino acid sequence of the
 CC human serine protease HIPHM46.
 XX
 SQ Sequence 882 AA;

Query Match 100.0%; Score 4700; DB 24; Length 882;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAAMETBQLGVEIFETADCEENISQDRPKLEPFVERYSQKLLADTRKHGYMM 60
 DB 1 MAAAMETBQLGVEIFETADCEENISQDRPKLEPFVERYSQKLLADTRKHGYMM 60
 QY 61 AKAHPDFMKNDPDGPHSDRIYYLAMSGENRENTLFPSEI PKTINRAAVLMSKPL 120
 DB 61 AKAHPDFMKNDPDGPHSDRIYYLAMSGENRENTLFPSEI PKTINRAAVLMSKPL 120
 QY 121 DLPQATLDYGMYSREBELRERKIGTVAGIADYHQSGLTLPQAGSIIYHVKDGPQ 180
 DB 121 DLPQATLDYGMYSREBELRERKIGTVAGIADYHQSGLTLPQAGSIIYHVKDGPQ 180
 QY 181 PFOQPLRPLVETSCPNIRMDPKLCPADPDWIAFHSNDIMVTSVTRERRLTYHNEL 240
 DB 181 PFOQPLRPLVETSCPNIRMDPKLCPADPDWIAFHSNDIMVTSVTRERRLTYHNEL 240
 QY 241 ANMEEDASAGATVLOEFPDYGYSYMMCPKATTPSGKILRIYENDESEVEIIHY 300
 DB 241 ANMEEDASAGATVLOEFPDYGYSYMMCPKATTPSGKILRIYENDESEVEIIHY 300
 QY 301 TSPMLETRRADSFYRYPKTGTANPKYTFKMSIEMIDAEGRIIDVIDKELIQPEILL 360
 DB 301 TSPMLETRRADSFYRYPKTGTANPKYTFKMSIEMIDAEGRIIDVIDKELIQPEILL 360
 QY 361 YIARAGMPEGKYAMSILLDRSOTLQIVLISPELFIPEDDVMERQRIIESVPSVTPL 420
 DB 361 YIARAGMPEGKYAMSILLDRSOTLQIVLISPELFIPEDDVMERQRIIESVPSVTPL 420
 QY 421 IYEEETDIIWINIHDIHFVFPQSHHEEIEFIPASECKTFRHLKYITSILKESKYRSSG 480
 DB 421 IYEEETDIIWINIHDIHFVFPQSHHEEIEFIPASECKTFRHLKYITSILKESKYRSSG 480
 QY 481 GLPAPSDPKCPKKEIATITSGMEVYLGRRHSNIQVDEVARLVFEGTKDPSPLEHLLYVVS 540
 DB 481 GLPAPSDPKCPKKEIATITSGMEVYLGRRHSNIQVDEVARLVFEGTKDPSPLEHLLYVVS 540
 QY 541 YNPGEVRLTRDRGSHSCCISQHCDFISKXSNKXNCHVSLYKLSPEDDPTCKTEF 600
 DB 541 YNPGEVRLTRDRGSHSCCISQHCDFISKXSNKXNCHVSLYKLSPEDDPTCKTEF 600
 QY 601 WATILDSAGPLPDYTPPEIFSFESTTGTLYGMLTKPHDLOPKKYPVLFIYGGPOVL 660
 DB 601 WATILDSAGPLPDYTPPEIFSFESTTGTLYGMLTKPHDLOPKKYPVLFIYGGPOVL 660
 QY 661 VNNRFGYKVFRLNLTASLGIVVVVITDNRGSGHRLKKEGAFKYMGGIIEIDQVEGLQY 720
 DB 661 VNNRFGYKVFRLNLTASLGIVVVVITDNRGSGHRLKKEGAFKYMGGIIEIDQVEGLQY 720
 QY 721 LASRYDFIDLRVGHGWSYGYLSLMAIMORSDFIRVALAGAPVTLMIFFDTGTERYM 780
 DB 721 LASRYDFIDLRVGHGWSYGYLSLMAIMORSDFIRVALAGAPVTLMIFFDTGTERYM 780
 QY 781 GHDDNQGCGYLLGSVMQAQKPESEPNRLLLHGFIDENHFAHTSILLSPVRAKGPYD 840

DB 781 GHPDNEQGYLLGSVAMQAEPSEPNRLLLHGFLENVHFAHTSILSLFLVAKGKPYD 840

QY 841 LQIYPOERHSIRVPESGEHYELHLHYLOENLGSRIALAKVI 882

DB 841 LQIYPOERHSIRVPESGEHYELHLHYLOENLGSRIALAKVI 882

RESULT 7
ABB97361
ID ABB97361 standard; Protein; 782 AA.
XX
AC ABB97361;
XX
XX 27-JUN-2002 (first entry)
XX
XX Novel human protein SEQ ID NO: 629.
XX

KM Human; antihaemic; vulnerary; antiinflammatory; immunomodulator;
KM antifertility; cerebroprotective; cytosolic; rheumatic; gene therapy;
KM neuroprotective; antiparkinsonian; protein therapy; EST;
KM expressed sequence tag.
XX
XX Homo sapiens.
XX
XX MO200222660-A2.
XX
XX 21-MAR-2002.
XX
XX 10-SEP-2001; 2001WO-US26015.
XX
XX 11-SEP-2000; 2000US-0659671.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX
XX Xue AJ, Yang Y, Wehrman T, Dermanac RT, Zhang QA, Ren F;
XX
XX WPI; 2002-292408/33.
XX
XX N-PSDB; ABB92547.
XX
XX

PT An isolated polynucleotide for treating diseases associated with its
XX
XX encoded polypeptide such as cancer and multiple sclerosis -
XX
XX
XX Example 2; SEQ ID NO 629; 509bp; English.
XX
XX

CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention.
XX
XX

Sequence 782 AA:
SQ

Query Match 87.6%; Score 4118; DB 23; Length 782;
Best Local Similarity 88.7%; Pred. No. 0;
Matches 782; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 MAAAMTBQLGVIFETADCEENIESODRPLKPPFYERYSWQLKLLADTKRYGYMM 60
DB 1 MAAAMTBQLGVIFETADCEENIESODRPLKPPFYERYSWQLKLLADTKRYGYMM 60
QY 61 AKAPHDPMFYKRDPPGPHSDRIYVLAMSGENRENTLFYSEIKTINRAAVLMSWPKLL 120
DB 61 AKAPHDPMFYKRDPPGPHSDRIYVLAMSGENRENTLFYSEIKTINRAAVLMSWPKLL 120
QY 121 DLFOATLDYGMYSREBELLRERKRIQTVGIASVDYHGSGTFLFOAGSGIYHVKGQGPQG 180
DB 121 DLFOATLDYGMYSREBELLRERKRIQTVGIASVDYHGSGTFLFOAGSGIYHVKGQGPQG 180

DB 121 DLFOATLDYGMYSREBELLRERKRIQTVGIASVDYHGSGTFLFOAGSGIYHVKGQGPQG 180

QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIYISNIVTSEERRLTVVHNEI 240

DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHNSNDIYISNIVTSEERRLTVVHNEI 240

QY 241 ANMEDARSAGVATFVLQEEFDRYSGYWMCCKAETTPSGGKILRLIYENDESEVEIIVH 300

DB 241 ANMEDARSAGVATFVLQEEFDRYSGYWMCCKAETTPSGGKILRLIYENDESEVEIIVH 300

QY 301 TSPMLFTRRADSFYRYPKTGTANPKYTRMSGIMIDABGRIIDVIDKELIQPEEILFEGVE 360

DB 301 TSPMLFTRRADSFYRYPKTGTANPKYTRMSGIMIDABGRIIDVIDKELIQPEEILFEGVE 360

QY 361 YIARAGWTPGKIVAMSTILLDRSOTRLQIVLISPELFIPEVDVNEROBLISVPSVPL 420

DB 361 YIARAGWTPGKIVAMSTILLDRSOTRLQIVLISPELFIPEVDVNEROBLISVPSVPL 420

QY 421 IIVEETDWINIHDIHFVFPQSHBEIEFIIPASECKTGFRHLKYKTSILKESKYKRSSG 480

DB 421 IIVEETDWINIHDIHFVFPQSHBEIEFIIPASECKTGFRHLKYKTSILKESKYKRSSG 480

QY 481 GLPAPSPFKCPKKEINAITSGEWEVLGRHGSNIQVDEVRLIYVEEGTKDSPLEHLLYVS 540

DB 481 GLPAPSPFKCPKKEINAITSGEWEVLGRHGSNIQVDEVRLIYVEEGTKDSPLEHLLYVS 540

QY 541 YVNGEYTRLRDRGYSHSCCISQCHDFEISKYKQKPHCVSLYKLSPEEDPTCKTKEF 600

DB 541 YVNGEYTRLRDRGYSHSCCISQCHDFEISKYKQKPHCVSLYKLSPEEDPTCKTKEF 600

QY 601 MATILDSAGPLPDYTPPELIFSFESETGFTLYGMLYKPHDLOPKKYPVLFIYGGPVQL 660

DB 601 MATILDSAGPLPDYTPPELIFSFESETGFTLYGMLYKPHDLOPKKYPVLFIYGGPVQL 660

QY 661 VNNRFKGVKIFRLNTLASLGVVVVIDNRGSCHRGLKPEGAFKYKMGQIEIDQVEGLQY 720

DB 661 VNNRFKGVKIFRLNTLASLGVVVVIDNRGSCHRGLKPEGAFKYKMGQIEIDQVEGLQY 720

QY 721 LASRYDFIDLDRAVIGHMSYGYLSLMAIMQRSDIFRVAIAGAPVTLMIPTDGYTERYM 780

DB 721 LASRYDFIDLDRAVIGHMSYGYLSLMAIMQRSDIFRVAIAGAPVTLMIPTDGYTERYM 780

QY 781 GHPDNEQGYLLGSVAMQAEPSEPNRLLLHGFLENVHFAHTSILSLFLVAKGKPYD 840

DB 781 GHPDNEQGYLLGSVAMQAEPSEPNRLLLHGFLENVHFAHTSILSLFLVAKGKPYD 840

QY 841 LQIYPOERHSIRVPESGEHYELHLHYLOENLGSRIALAKVI 882

DB 841 LQIYPOERHSIRVPESGEHYELHLHYLOENLGSRIALAKVI 882

RESULT 8
ABB97362
ID ABB97362 standard; Protein; 724 AA.
XX
AC ABB97362;
XX
XX 27-JUN-2002 (first entry)
XX
XX Novel human protein SEQ ID NO: 630.
XX
XX Human; antihaemic; vulnerary; antiinflammatory; immunomodulator;
XX
XX KM antifertility; cerebroprotective; cytosolic; rheumatic; gene therapy;
XX
XX KM neuroprotective; antiparkinsonian; protein therapy; EST;
XX
XX expressed sequence tag.
XX
XX Homo sapiens.
XX
XX MO200222660-A2.
XX
XX 21-MAR-2002.
XX
XX 10-SEP-2001; 2001WO-US26015.
XX

```

XX 11-SEP-2000; 2000US-0659671.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F,
PI Xue AU, Yang Y, Wehrman T, Dimanac RT;
XX
XX WPI; 2002-292408/33.
DR N-PSDB; AEN32548.
XX
PT An isolated polynucleotide for treating diseases associated with its
FT encoded polypeptide such as cancer and multiple sclerosis -
XX
XX Example 2; SEQ ID NO 630; 509pp; English.
PS
XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention.
XX
SQ Sequence 724 AA;
Query Match 80.2%; Score 3771; DB 23; Length 724;
Best Local Similarity 82.1%; Pred. No. 0;
Matches 724; Conservative 0; Mismatches 0; Indels 158; Gaps 2;
QY 1 MAAAMEETOLGVELEFETADCEENIESQDRPLPEFYERYYSQKLKLDTRKHYGYM 60
DB 1 MAAAMEEQLGVELEFETADCEENIESQDRPLPEFYERYYSQKLKLDTRKHYGYM 60
QY 61 AKAPHDPMFYKRNDDGPHSDRIYYLMSGENRENTLFYSEIPIKTINRAAVLMSKPL 120
DB 61 AKAPHDPMFYKRNDDGPHSDRIYYLMSGENRENTLFYSEIPIKTINRAAVLMSKPL 120
QY 121 DLPGATIDYGMVSEEEELRERKIGTVGIASVYHGGSGFLFQAQSGIYHVDGPGG 180
DB 121 DLPG----- 124
QY 181 FTQGPLRPNI VETS CPNIRMDPKLCPADPMIAFIHSNDIWSNI VTRERRLLTYVANEL 240
DB 125 --QQPLRPNI VETS CPNIRMDPKLCPADPMIAFIHSNDIWSNI VTRERRLLTYVANEL 182
QY 241 ANMEDARSAGVATFVYOEFDRYSGYWCPCAKETTPSGKILRLIYEENDESVEIIVH 300
DB 183 ANMEDARSAGVATFVYOEFDRYSGYWCPCAKETTPSGKILRLIYEENDESVEIIVH 242
QY 301 TSPMLFTRRADSFPYPTGTANPKVTFKMSIIMDAGRIIDVIDKELIQPELLEFGVE 360
DB 243 TSPMLFTRRADSFPYPTGTANPKVTFKMSIIMDAGRIIDVIDKELIQPELLEFGVE 302
QY 361 YIARAGTPEGKYAMSIILDRSQTRLQIVLISPFLIPVEDDWERORLIESVDSYTPL 420
DB 303 YIARAGTPEGKYAMSIILDRSQTRLQIVLISPFLIPVEDDWERORLIESVDSYTPL 362
QY 421 IIVEETDITNINHDIHFVFPQSHHEEIEFIASECTGGRHLKYTISLKESYKXSSG 480
DB 363 IIVEETDITNINHDIHFVFPQSHHEEIEFIASECTGGRHLKYTISLKESYKXSSG 422
QY 481 GLPAPSDFKPIKEEIAITSGEMVELGRHGSNIQVDEVRLLVPEGTQKSPLEHHLVYS 540
DB 423 GLPAPSDFKPIKEEIAITSGEMVELGRHGSNIQVDEVRLLVPEGTQKSPLEHHLVYS 482
QY 541 YVNPGEVTRLLDRGYSHSCCISQHCDFPISKYSNQNPKCVSLYKLSPPDDPTCKTKEF 600
DB 483 YVNPGEVTRLLDRGYSHSCCISQHCDFPISKYSNQNPKCVSLYKLSPPDDPTCKTKEF 542

```

```

QY 601 MATLDSAGLPLDPTPEIIFSFSTTGTLLYGMLYKPHDQPGKPYTVLFIYGGPQVL 660
DB 543 MATLDSAGLPLDPTPEIIFSFSTTGTLLYGMLYKPHDQPGKPYTVLFIYGGPQ--- 599
QY 661 VNNRFKGVKFRNLNTASLGIVVVVIDNRGSCHRGLKFBGAFKXRMQGLEIDQVGLQY 720
DB 600 ----- 599
QY 721 LASKYDIDLDVRGSHWMSVGYLSIMALMQRSDIFPVATAGAPVTLMIFPDGYTERYM 780
DB 600 -----VALAGAPVTLMIFPDGYTERYM 622
QY 781 GHPDQNGGYVLSVAMQAEKPEPRLLLHGFIDENVAFATSIILSLFVRAKRPYD 840
DB 623 GHPDQNGGYVLSVAMQAEKPEPRLLLHGFIDENVAFATSIILSLFVRAKRPYD 682
QY 841 LQIYPOERHSIRVSEGEHYELHLYLOENLGSRIALKVI 882
DB 683 LQIYPOERHSIRVSEGEHYELHLYLOENLGSRIALKVI 724

```

RESULT 9

ID ABG61600 standard; Protein: 658 AA.

XX ABG61600;

DT 12-AUG-2002 (first entry)

XX Human DPP-1 splice variant #7.

XX Human, serine protease; dipeptidyl peptidase IV-related protein; DPP; DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain; diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke; heart failure; hypertension; urinary retention; osteoporosis; cancer; ulcer; allergy; cancer; psychotic disorder; neurological disorder; dyskinnesia; reproductive disorder; inflammatory disorder; metabolic disorder.

XX Homo sapiens.

XX W0200231134-A2.

XX 18-APR-2002.

XX 12-OCT-2001; 2001WO-US31874.

XX 12-OCT-2000; 2000US-240117P.

XX (FERR) FERRING BV.

XX Q1 S, Akinsanya KO, Riviere PJ, Junien J;

XX WPI; 2002-444178/47.

XX N-PSDB; ABK83331.

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding the proteins, useful for treating e.g. fungal, bacterial, protozoan and viral infections, cancers, allergies, neurological disorders, or pain

XX Disclosure; Page 70-72; 113pp; English.

XX The present invention relates to the isolation of novel human serine proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP) and nucleic acids encoding them are useful for treating infections such as fungal, bacterial, protozoan and viral infections, particularly infections caused by human immunodeficiency virus (HIV-1 or HIV-2), CC pain, diabetes, precocious puberty, infertility, obesity, anorexia, CC bulimia, Parkinson's disease, acute heart failure, hypotension, CC hypertension, urinary retention, osteoporosis, angina pectoris, CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,

CC psychotropic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 658 AA;

Query Match 74.6%; Score 3504; DB 23; Length 658;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAAAMETQQLGVEIFETADCEINIESODRPLKEPFYVERYSQKLLADTRKHGYMM 60
DB 1 MAAAMETQQLGVEIFETADCEINIESODRPLKEPFYVERYSQKLLADTRKHGYMM 60
QY 61 AKAPHEFVFRKNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTKINRAAVLMSKPELL 120
DB 61 AKAPHEFVFRKNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTKINRAAVLMSKPELL 120
QY 121 DLFOATLDYGMYSREBELLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXKDGPGQ 180
DB 121 DLFOATLDYGMYSREBELLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXKDGPGQ 180
QY 121 DLFOATLDYGMYSREBELLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXKDGPGQ 180
DB 121 DLFOATLDYGMYSREBELLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXKDGPGQ 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDIMISNIVTREERLLTYVHNEL 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDIMISNIVTREERLLTYVHNEL 240
QY 241 ANMEDARSAGVATFVLOEEDFRYSGYWMCCKAETTPSGKILRLIYEENDESEVEIHHV 300
DB 241 ANMEDARSAGVATFVLOEEDFRYSGYWMCCKAETTPSGKILRLIYEENDESEVEIHHV 300
QY 301 TSPMLERTRADSFRYPKGTANPKVTFKMSIEMIDAEGRIIDVIDKELIQPEILFEGVE 360
DB 301 TSPMLERTRADSFRYPKGTANPKVTFKMSIEMIDAEGRIIDVIDKELIQPEILFEGVE 360
QY 361 YIARAGWTPBEKRYAWSILLDRSQRLQIVLISPFLFVDDVWERQRLISVDSVTPPL 420
DB 361 YIARAGWTPBEKRYAWSILLDRSQRLQIVLISPFLFVDDVWERQRLISVDSVTPPL 420
QY 421 IIVETTDIMINIHDIHFVFPQSHHEIEFTFASECKTFPHLYKITSILKESKYSRSG 480
DB 421 IIVETTDIMINIHDIHFVFPQSHHEIEFTFASECKTFPHLYKITSILKESKYSRSG 480
QY 481 GLPAPSDPKPIKEEIAITSGBEVYLGHSGNIQVDEVRLVYFEGTKOSPLEHNLVYVS 540
DB 481 GLPAPSDPKPIKEEIAITSGBEVYLGHSGNIQVDEVRLVYFEGTKOSPLEHNLVYVS 540
QY 541 YVNGEVTURLTDRGSHSCCISOHCDPFIISKYSNOKNHCVSILKLSPEDDPTCKTEF 600
DB 541 YVNGEVTURLTDRGSHSCCISOHCDPFIISKYSNOKNHCVSILKLSPEDDPTCKTEF 600
QY 601 WATILDSAGPLPDYTPPEIFSESTTGTLYGMLYKPHDLQPKKKYPTVLEFYGG 655
DB 601 WATILDSAGPLPDYTPPEIFSESTTGTLYGMLYKPHDLQPKKKYPTVLEFYGG 655

```

RESULT 10
 ABG61596
 ID ABG61596 standard; Protein; 661 AA.
 XX
 AC ABG61596;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPRP-1 splice variant #3.
 XX

KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder.

XX Homo sapiens.
 OS
 XX
 PN W0200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX

XX (FEER) FERRING BV.
 PA
 XX Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX WPI; 2002-444178/47.
 DR N-P8DB; ARK83327.
 XX

XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 XX

PS Disclosure; Page 63-65; 113pp; English.

CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
 CC

SQ Sequence 661 AA;

Query Match 74.6%; Score 3504; DB 23; Length 661;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAAAMETQQLGVEIFETADCEINIESODRPLKEPFYVERYSQKLLADTRKHGYMM 60
DB 1 MAAAMETQQLGVEIFETADCEINIESODRPLKEPFYVERYSQKLLADTRKHGYMM 60
QY 61 AKAPHEFVFRKNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTKINRAAVLMSKPELL 120
DB 61 AKAPHEFVFRKNDPDGPHSDRIYYLAMSNGENRENTLYSEIPTKINRAAVLMSKPELL 120
QY 121 DLFOATLDYGMYSREBELLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXKDGPGQ 180
DB 121 DLFOATLDYGMYSREBELLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXKDGPGQ 180
QY 121 DLFOATLDYGMYSREBELLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXKDGPGQ 180
DB 121 DLFOATLDYGMYSREBELLRERKRIIGTVGLASVDYHQSSTFLFQAGSGIYHVXKDGPGQ 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDIMISNIVTREERLLTYVHNEL 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDIMISNIVTREERLLTYVHNEL 240
QY 241 ANMEDARSAGVATFVLOEEDFRYSGYWMCCKAETTPSGKILRLIYEENDESEVEIHHV 300
DB 241 ANMEDARSAGVATFVLOEEDFRYSGYWMCCKAETTPSGKILRLIYEENDESEVEIHHV 300
QY 301 TSPMLERTRADSFRYPKGTANPKVTFKMSIEMIDAEGRIIDVIDKELIQPEILFEGVE 360
DB 301 TSPMLERTRADSFRYPKGTANPKVTFKMSIEMIDAEGRIIDVIDKELIQPEILFEGVE 360
QY 361 YIARAGWTPBEKRYAWSILLDRSQRLQIVLISPFLFVDDVWERQRLISVDSVTPPL 420
DB 361 YIARAGWTPBEKRYAWSILLDRSQRLQIVLISPFLFVDDVWERQRLISVDSVTPPL 420

```

QY 421 IYEEETDWMINIHDFHVPFQSHHEIEIFAFSECKTGFRHLKYKITSILKSKYKRS6G 480
 DB 421 IYEEETDWMINIHDFHVPFQSHHEIEIFAFSECKTGFRHLKYKITSILKSKYKRS6G 480
 QY 481 GLPAPSDFKCPKEIEAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVS 540
 DB 481 GLPAPSDFKCPKEIEAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVS 540
 QY 541 YVNPGEVTRLTDRGSHSCCISOHCDPFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600
 DB 541 YVNPGEVTRLTDRGSHSCCISOHCDPFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600
 QY 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGG 655
 DB 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGG 655

RESULT 11

ID ABG61594 standard; Protein; 690 AA.

AC ABG61594;

DT 12-AUG-2002 (first entry)

DE Human DPRP-1 splice variant #1.

KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KM DPPIV; infection; human immunodeficiency virus; HIV-1, HIV-2, pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinetic; reproductive disorder; inflammatory disorder;
 KM metabolic disorder.

OS Homo sapiens.

FN WO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001MO-US31874.

PR 12-OCT-2000; 2000US-240117P.

PA (FERR) FERRING BV.

PI Qi S, Akinsanya KO, Riviere PJ, Junien J;

DR MPI, 2002-444178/47.

DR N-PSDB; ABR83325.

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain

PS Disclosure; Page 59-61; 113pp; English.

CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and

CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 690 AA:

Query Match 74.6%; Score 3504; DB 23; Length 690;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAAMETQOLGVEIFETADCEENIESODRPKLEPEYVERYSQKLLADTRKHYGMM 60

DB 1 MAAAMETQOLGVEIFETADCEENIESODRPKLEPEYVERYSQKLLADTRKHYGMM 60

QY 61 AKAPHDFFVKNRNDPGEHSDRIYVLAAMSGENRENTLFYSELPKINRAAVMLGKPL 120

DB 61 AKAPHDFFVKNRNDPGEHSDRIYVLAAMSGENRENTLFYSELPKINRAAVMLGKPL 120

QY 121 DLFQATLDYGMYSRREBELRERKIGTVGIASVDYHQSGLFLPQAGSGIYHVKGQPG 180

DB 121 DLFQATLDYGMYSRREBELRERKIGTVGIASVDYHQSGLFLPQAGSGIYHVKGQPG 180

QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNVTREERLTVVHNL 240

DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNVTREERLTVVHNL 240

QY 241 ANMEDARSAGVATFVLQEEFDRYSGYWMCFAETTPSGKILRLYEENDESEYIIV 300

DB 241 ANMEDARSAGVATFVLQEEFDRYSGYWMCFAETTPSGKILRLYEENDESEYIIV 300

QY 301 TSPMLETFRASFRPKGTANPKTFKSGEIMIDABRIIDVILKEILOPELIFEGVE 360

DB 301 TSPMLETFRASFRPKGTANPKTFKSGEIMIDABRIIDVILKEILOPELIFEGVE 360

QY 361 YIARAGWPEEGVAMSIILDRSQRLQVILSPELFIVEDVMEORLIEVPSVPL 420

DB 361 YIARAGWPEEGVAMSIILDRSQRLQVILSPELFIVEDVMEORLIEVPSVPL 420

QY 421 IYEEETDWMINIHDFHVPFQSHHEIEIFAFSECKTGFRHLKYKITSILKSKYKRS6G 480

DB 421 IYEEETDWMINIHDFHVPFQSHHEIEIFAFSECKTGFRHLKYKITSILKSKYKRS6G 480

QY 481 GLPAPSDFKCPKEIEAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVS 540

DB 481 GLPAPSDFKCPKEIEAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVS 540

QY 541 YVNPGEVTRLTDRGSHSCCISOHCDPFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600

DB 541 YVNPGEVTRLTDRGSHSCCISOHCDPFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600

QY 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGG 655

DB 601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGG 655

RESULT 12

ABG93565

ID AAB93565 standard; Protein; 632 AA.

AC AAB93565;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:12964.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUN-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T,
 XX
 DR WPI; 2001-318749/34.
 XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX

Claim 8; SEQ ID 12964; 2537bp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides, and the combination
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB58583 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC
 XX
 XX

Sequence 632 AA;

Query Match 70.9%; Score 3333.5; DB 22; Length 632;
 Best Local Similarity 92.4%; Pred. No. 1.7e-312;
 Matches 631; Conservative 0; Mismatches 1; Indels 51; Gaps 1;

QY 200 MDPKLCPPADPDMIAFIHSNDIWMISNIVTRERRLTYVHNELANMEDARSAGVATFVIOE 259
 DB 1 MDPKLCPPADPDMIAFIHSNDIWMISNIVTRERRLTYVHNELANMEDARSAGVATFVIOE 60
 QY 260 EFDYSGYWCPCAKETTPSGKILRLIYENDSEVEIHTSPLMLETERRADSPFPYKTG 319
 DB 61 EFDYSGYWCPCAKETTPSGKILRLIYENDSEVEIHTSPLMLETERRADSPFPYKTG 120
 QY 320 TANPKVTFKXSEIMIDAGRIIDVYDKELIOPFEILFEGVYIARAGTPEKGYAWMSILL 379
 DB 121 TANPKVTFKXSEIMIDAGRIIDVYDKELIOPFEILFEGVYIARAGTPEKGYAWMSILL 180
 QY 380 DRSGTRLOIVLISELFIPEDDVWERORLIESVPDSVTPILTYEETTDIWINIHDPHV 439
 DB 181 DRSGTRLOIVLISELFIPEDDVWERORLIESVPDSVTPILTYEETTDIWINIHDPHV 240
 QY 440 FPQSHHEIEIEIFASECKTGFRHLKYITSLIKESKYKSSGGLPAPDFKCPKEEIAIT 499
 DB 241 FPQSHHEIEIEIFASECKTGFRHLKYITSLIKESKYKSSGGLPAPDFKCPKEEIAIT 300
 QY 500 SGEWVUGRGSNIOVDEVRRLVYFEGTKOSPLFHHLYVVSYNVPGVTRITDRGYSHSC 559
 DB 301 SGEWVUGRGSNIOVDEVRRLVYFEGTKOSPLFHHLYVVSYNVPGVTRITDRGYSHSC 360

QY 560 CISQHCDFPISKYSNOKNPHCVSLYKLSSPEDDPCTCKEFAFATILDSAGPLDPTYPEI 619
 DB 361 CISQHCDFPISKYSNOKNPHCVSLYKLSSPEDDPCTCKEFAFATILDSAGPLDPTYPEI 420
 QY 620 FSPFSTTGTLVYMLYKPHDLOPKKYPTVLFTYGGPQVQVNNRRKGVYFELNLTASL 679
 DB 421 FSPFSTTGTLVYMLYKPHDLOPKKYPTVLFTYGGPQVQVNNRRKGVYFELNLTASL 480
 QY 680 GYVVVVVIDRSGSCHRGKKEGAFKYMGGQIEIDQVEGQVYLA SRDFIDLRVGIHWS 739
 DB 481 GYVVVVVIDRSGSCHRGKKEGAFKYMGGQIEIDQVEGQVYLA SRDFIDLRVGIHWS 507
 QY 740 YGGYLSLMAIMORSDIFRVAIAGAPVTLWIFDGTGYTERYMGHPDQEGYLSYVAMQA 799
 DB 508 YGGYLSLMAIMORSDIFRVAIAGAPVTLWIFDGTGYTERYMGHPDQEGYLSYVAMQA 549
 QY 800 EKPSEPNRLLHGLDENVHFAHVSILSLFVRAKRYDQIYQERHSIRVPESGEH 859
 DB 550 EKPSEPNRLLHGLDENVHFAHVSILSLFVRAKRYDQIYQERHSIRVPESGEH 609
 QY 860 YELHLHLHYLQENLGSRIALAKVI 882
 DB 610 YELHLHLHYLQENLGSRIALAKVI 632

RESULT 13
 AAB61601
 ID AAB61601 strand; Protein: 613 AA.
 XX
 AC AAB61601;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPRP-1 splice variant #8.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPP;
 KW DPPP; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinetic; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 KW
 OS Homo sapiens.
 XX
 XX WO20021134-A2.
 XX
 XX 18-APR-2002.
 PD
 XX 12-OCT-2001; 2001WO-US31874.
 PF
 XX 12-OCT-2000; 2000US-240117P.
 PR
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Jansen J;
 XX
 DR WPI; 2002-444178/47.
 XX
 DB N-PSDB; ABE63332.
 XX
 XX

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 PS Disclosure; Page 73-75; 113pp; English.
 XX
 XX

CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPP) -related
 CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC
 CC

CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinestias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

SQ Sequence 613 AA;

Query Match 68.9%; Score 3236; DB 23; Length 613;

Best Local Similarity 100.0%; Pred. No. 4, 2e-303; Mismatches 607; Conservative 0; Indels 0; Gaps 0;

1 MAAMETEQLGVEIFFETADCEENIESQDRPKLEPFYVERYSQKLADLRKHYGM 60
1 MAAMETEQLGVEIFFETADCEENIESQDRPKLEPFYVERYSQKLADLRKHYGM 60
61 AQAHDPMFVKRNDPDGPHSDRIYVLAAMGENRENTLFYSEIPKTIINRAVIMLSWKPL 120
61 AQAHDPMFVKRNDPDGPHSDRIYVLAAMGENRENTLFYSEIPKTIINRAVIMLSWKPL 120
121 DLFQATLDYGYMSREBELRERKRTIGVGIASYDHQSGTFLFQSGGIYHVKOGPQG 180
121 DLFQATLDYGYMSREBELRERKRTIGVGIASYDHQSGTFLFQSGGIYHVKOGPQG 180
181 FTQOQLRPNLVETSCENIMDPKLCPADPDWIAFIHNSDIWISNTYRERRLTYVHNL 240
181 FTQOQLRPNLVETSCENIMDPKLCPADPDWIAFIHNSDIWISNTYRERRLTYVHNL 240
241 AMMEEDARASGAVATFLQSEFDRISGYMWCPRKATTPSGGKILRIIYENDSEVEI 300
241 AMMEEDARASGAVATFLQSEFDRISGYMWCPRKATTPSGGKILRIIYENDSEVEI 300
301 TSPMLETRADSEFRYKGTANPKVTKMSEIMIAEGRIIDIVIKELIQPELIFEGVE 360
301 TSPMLETRADSEFRYKGTANPKVTKMSEIMIAEGRIIDIVIKELIQPELIFEGVE 360
361 YIARAQWTEPEGKYAMSIILDRSQTRLQIVLISPELFIPEDDVMERQRLIESVDP 420
361 YIARAQWTEPEGKYAMSIILDRSQTRLQIVLISPELFIPEDDVMERQRLIESVDP 420
421 IYETETDITWNIHIDIFHVFPOSHHEEIFFIPASECKTGFRHLKYITSLKSKYKRS 480
421 IYETETDITWNIHIDIFHVFPOSHHEEIFFIPASECKTGFRHLKYITSLKSKYKRS 480
481 GLPAPSDFCPIKEELIAITSGEMEVLAGHGSNIQYDEVRLVYFEGTQXSPLEHLLYVS 540
481 GLPAPSDFCPIKEELIAITSGEMEVLAGHGSNIQYDEVRLVYFEGTQXSPLEHLLYVS 540
541 YNPGEVRLRLDRGYSHSCCISQHCDFISKYSNOKNPHCVSLYKLSBEDDPTCKTEF 600
541 YNPGEVRLRLDRGYSHSCCISQHCDFISKYSNOKNPHCVSLYKLSBEDDPTCKTEF 600
601 WATILDS 607
601 WATILDS 607

RESULT 14
ID ABG61592 standard; Protein; 863 AA.
AC ABG61592;
DT 12-AUG-2002 (first entry)
DE Human DPRP related serine protease DPRP-2.
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;

KM heart failure, hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyskinestias; reproductive disorder; inflammatory disorder;
KM metabolic disorder.

OS Homo sapiens.

PN MO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

PR 12-OCT-2000; 2000US-240117P.

PA (FERR) FERRING BV.

PI Qi S, Akinsanya KO, Riviere PJ, Junien J;

DR WPI; 2002-444178/47.

DR N-PSDB; ABK63323.

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain

Claim 17; Fig 1; 113pp; English.

CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinestias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

SQ Sequence 863 AA;

Query Match 61.1%; Score 2870; DB 23; Length 863;

Best Local Similarity 61.5%; Pred. No. 1.8e-267; Mismatches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;

35 FVYERYSQKLADLRKHYGMKAPADPMFVKRNDPDGPHSDRIYVLAAMGENE 94
24 FVQKQSHNDGLRSTIHSRKYSGLIVKAPADPMFVKRNDPDGPHSDRIYVLAAMGENE 83
95 NTLFYSEIPKTIINRAVIMLSWKPLDLFOATLDYGYMSREBELRERKRTIGVGIASYD 154
84 NSLIYSEIPKTYRKEALLLSWKQMLDHFQATPHNGVSRBELRERKRLGVFETITD 143
155 YHOGSGTFLFOAGSGIYHVKOGPQGFQOQLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
144 FHSBSGLFLFOASNSLFCRCDGKNGFVSPMKYLEIKTQCSGPRMDKICPADAPFSPF 203
215 IHSNDIWSNTVTRERRLTYVHNELANMEEDARASGAVATVLOEFPYRSGYMWCPRAE 274
204 INNSDIWAMANTETBERRLTCHQGLSNVLDPPASAGVATVTOEFPRTGYMWCPYAS 263
275 TTPSGG-KILRIIYENDSEVEIIVHTSPMLETRADSEFRYKGTANPKVTKMSEIM 333
264 WEGSEGGKTLRIILYEVDESEVEIIVHSPALERKTDYVYPTGSGNPIALKLAFFQ 323
336 IDARGRIIDVIDEKLQPELIFEGVYIYIARAGTTPGKYAMSIILDRSQTRLQIVLISP 393
324 TDSQGIKIVTQEKELVQFSSLFEPKVEYIARAAGWTRDGKYAMAMFLDRPQOQLVILPP 383

Db 533 I W N E E T K L V Y F O G I K O T P L E H H L V V S Y E A G E I V R L T P G F S H S C M S Q N F D M F V S H Y 592
QY 573 S N O K N P C V S L Y K L S S P E D D P T C K T E F W A T I L D S A G E L P Y T P E I F S E S T G F T L Y G 632
Db 593 S S V S T P C V A V Y K L S G P D D P L H K O P R F W A S M E A S C P D Y P P E I F H F H T R S D V R L Y G 652
QY 633 M L Y K P H D L O P G K K Y P T V L F I Y G P O V O L V N N R F K G V K Y F R L N T L A S L G Y V V V I D N R G S C 692
Db 653 M I Y K P H A L O P G K K H P T V L F V Y G P O V O L V N N S P F K I K Y L R L N T L A S L G Y A V V I D G R G S C 712
QY 693 H R G I K F E G A F Y K M G O I R I D D O V E G L O Y L A S R Y D F I D L D R V G I H G M S Y G Y L S I M A L M O R 752
Db 713 Q R G L R F E G A L K N O M G O V E I E D O V E G L O F V A E K Y F I D L S R V A I H G M S Y G F L S I M G L I H K 772
QY 753 S D I F R V A I A G A P V T L M I F Y D T G Y T E R Y W G H P D N O G Y L G S V A M O A E K P S E P N R L L L 812
Db 773 P O V F K V A I A G A P V T V M A Y D I G Y T E R M D V P E N N O H G E A G S V A L H V E K L P N E P N R L L I L 832
QY 813 H G F L D E N V H F A H T S I L L S F L V R A G K P Y D L O I Y P O E R H S I R V P E S G E H Y E L H L A Y L O E N L 872
Db 833 H G F L D E N V H F E H T N F L V S O L I R A G K P Y O L O I Y P N E R H S I R C P E S G E H Y E V T L I H F L O E Y L 892

Search completed: October 15, 2003, 17:10:43
Job time : 58.7856 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:10:56 ; Search time 19.6777 Seconds
(without alignments)
1896.467 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700
Sequence: 1 MAAMETEQLGVEIFETADC.....HLHYLQENLSRIALKVI 882

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents AA.*
2: /cgn2_6/ptodata/1/iaa/5A COMB.pdp.*
3: /cgn2_6/ptodata/1/iaa/5B COMB.pdp.*
4: /cgn2_6/ptodata/1/iaa/6A COMB.pdp.*
5: /cgn2_6/ptodata/1/iaa/6B COMB.pdp.*
6: /cgn2_6/ptodata/1/iaa/PTUS COMB.pdp.*
7: /cgn2_6/ptodata/1/iaa/backfile1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1845.5	35.0	310	4	US-09-794-236-4
2	582.5	12.4	771	4	US-09-462-284-2
3	529	11.3	766	4	US-10-002-593-6
4	525	11.2	755	5	PCT-US93-07923-3
5	525	11.2	759	5	PCT-US93-07923-2
6	525	11.2	766	1	US-08-230-491A-3
7	525	11.2	766	1	US-08-619-280A-3
8	525	11.2	766	2	US-08-940-391-3
9	525	11.2	766	4	US-09-794-236-1
10	454	9.7	760	1	US-08-230-491A-2
11	454	9.7	760	1	US-08-619-280A-2
12	454	9.7	760	2	US-08-940-391-2
13	282	6.0	632	3	US-09-016-080-1
14	249.5	5.3	657	4	US-09-355-166-1
15	225.5	4.8	593	5	PCT-US93-07923-11
16	192	4.1	622	2	US-08-664-46A-2
17	192	4.1	622	2	US-09-066-285-2
18	192	4.1	622	3	US-09-261-006-2
19	182	4.1	622	3	US-08-951-088-2
20	182	4.1	622	3	US-09-609-566-2
21	192	4.1	622	4	US-09-609-570-2
22	192	4.1	622	4	US-09-427-372-2
23	192	4.1	622	4	US-09-693-554-2
24	180.5	3.8	622	4	US-09-390-234-20
25	175.5	3.7	614	4	US-09-252-991A-20060
26	133	2.8	344	4	US-09-724-623-79
27	132.5	2.8	275	4	US-09-198-452A-181

28	123	2.6	1022	1	US-08-271-364A-8	Sequence 8, Appl1
29	123	2.6	1022	2	US-08-222-715B-27	Sequence 27, Appl1
30	120.5	2.5	686	6	US-09-368-169-8	Sequence 8, Appl1
31	116	2.5	561	3	US-09-233-989-5	Sequence 5, Appl1
32	113.5	2.4	1209	4	US-09-749-588-4	Sequence 4, Appl1
33	112.5	2.4	529	4	US-09-252-991A-24711	Sequence 24711, A
34	112.5	2.4	1285	1	US-07-582-945-2	Sequence 2, Appl1
35	112.5	2.4	1285	2	US-08-453-141-2	Sequence 2, Appl1
36	112.5	2.4	1285	3	US-08-293-314-2	Sequence 2, Appl1
37	110.5	2.4	836	4	US-09-491-356C-21	Sequence 21, Appl1
38	110	2.3	1296	1	US-08-480-604A-28	Sequence 28, Appl1
39	110	2.3	1296	2	US-08-405-496A-28	Sequence 28, Appl1
40	110	2.3	1296	3	US-08-915-136-28	Sequence 28, Appl1
41	109	2.3	570	3	US-09-068-960-13	Sequence 13, Appl1
42	109	2.3	808	4	US-09-198-452A-327	Sequence 327, App
43	108.5	2.3	782	4	US-09-585-858-29	Sequence 29, Appl1
44	106	2.3	256	4	US-09-355-166-2	Sequence 2, Appl1
45	106	2.3	1362	4	US-09-134-001C-4955	Sequence 4955, Ap

ALIGNMENTS

RESULT 1
US-09-794-236-4
Sequence 4, Application US/09794236
Patent No. 6337069
GENERAL INFORMATION:
APPLICANT: Grouzmann, Eric
APPLICANT: Lacroix, Jean-Silvain
TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
FILE REFERENCE: 81985/276823
CURRENT APPLICATION NUMBER: US/09/794, 236
CURRENT FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 310
TYPE: PRT
ORGANISM: Homo sapiens
US-09-794-236-4

Query Match	Best Local Similarity	Score	DB 4	Length	310
Matches 310; Conservative	0; Mismatches	0; Indels	49; Gaps	1;	
QY 524	PEGTKSPLEHHLVYVSYNPGVYRLTDRGYSHSCISQHCDFEISKYSNOKNPHCVSL	583			
DB 1	FEETKSPLEHHLVYVSYNPGVYRLTDRGYSHSCISQHCDFEISKYSNOKNPHCVSL	60			
QY 584	YKLSPPDDPTCKTKERFATILDSAGPLDPTTPEIFSEFSTTGTFLYGMLYKPHDLOPG	643			
DB 61	YKLSPPDDPTCKTKERFATILDSAGPLDPTTPEIFSEFSTTGTFLYGMLYKPHDLOPG	120			
QY 644	KKYPVLYFIYGPVOVLVNNRFKGVYKFRMLNTLASLIGYVVVINDRSGHGLKEGAFK	703			
DB 121	KKYPVLYFIYGPVOVLVNNRFKGVYKFRMLNTLASLIGYVVVINDRSGHGLKEGAFK	134			
QY 704	YKRGQLEIDQVQGLQYLASRYDFIDLDRVINGWSYGYGLSLMALMORSDFRVALAGA	763			
DB 135	--GQLEIDQVQGLQYLASRYDFIDLDRVINGWSYGYGLSLMALMORSDFRVALAGA	191			
QY 764	PVTLMTFYDTGYERYMGHPDQEGYVLSVAMQEKPSSENRLLHLHGFIDENVHFA	823			
DB 192	PVTLMTFYDTGYERYMGHPDQEGYVLSVAMQEKPSSENRLLHLHGFIDENVHFA	251			
QY 824	HTSILSLFVRAKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENLSRIALKVI	882			
DB 252	HTSILSLFVRAKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENLSRIALKVI	310			

RESULT 2

US-09-462-284-2
 ; Sequence 2, Application US/09462284
 ; Patent No. 6309868
 ; GENERAL INFORMATION:
 ; APPLICANT: Nestec S.A.
 ; APPLICANT: Monod, Michel
 ; APPLICANT: Donnas, Agnes
 ; APPLICANT: Affolter, Michael
 ; APPLICANT: Van Den Broek, Peter
 ; TITLE OF INVENTION: CLONING OF THE
 ; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
 ; TITLE OF INVENTION: ASPERGILLUS ORYZAE
 ; FILE REFERENCE: 8265-298
 ; CURRENT APPLICATION NUMBER: US/09/462,284
 ; CURRENT FILING DATE: 2000-01-03
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 771
 ; TYPE: PRT
 ; ORGANISM: Fungus
 ; US-09-462-284-2

Query Match 12.4%; Score 582.5; DB 4; Length 771;
 Best Local Similarity 28.4%; Pred. No. 7,66-50;
 Matches 198; Conservative 103; Mismatches 262; Indels 135; Gaps 30;
 QY 203 KLCRADP-----W-----IAFHSND--IWISNYTREERRLLTYVHNELANNEE 245
 DB 136 KSVPLMPQESDIOYAQMSPVGNLTAFVRENDLYVWDNGTVTR-----ITD 181
 QY 246 DA---RSAGVATFVLQEEF--DRYGYWMCRAETTPSGKILRLIYEENDESEVEIITHV 300
 DB 182 DGGDFMFGVDPMDIYEEIILGDRYA-LMFSPDGE-----YIAYISFNETGPTVTV 231
 QY 301 TSPMLERRADSF-----RYPKGTGANPKVTFKXSEIMIDAEGRIIDIVIKELIQPEI 354
 DB 232 QYVMDNOEIAIAPYPMELKIRYPKVSQTNPTVTLs-----LNIATAKEVQAID 280
 QY 355 LFESEVYIARAGMPEKGVYMSIILDRSQTLQVILSPFLPIYEDDVMERQRLIESVP 414
 DB 281 AFESTDLIT-----GEVAM--LTDHTT-----VAKAFNRVOD--QOKVAVADTAS 323
 QY 415 DSVPLIYEETDIWI-NIHDIHFVFPQSHHEIEFIASECKTGPHLYKISILKES 473
 DB 324 NKAT-VISDRGTGMDNLMSKYGIRKSDKDAYITIDISGMAHLTFP----- 376
 QY 474 KYKSSGGLPAPSDPKPIKEEIAITSGEMEVILGRHSNIQVDEVRIIVYFEGTKDSPLE 533
 DB 377 ---VSGGEPIP-----LTKGDMEVT---SILSDQERQVLYVLSQHHSTE 416
 QY 534 HHLVYVYVNPGEVTRITDRGYS--HSCCISQHCDFPIKSKSNOKNHCVSILKSPED 591
 DB 417 RHLXSVY-STFAVTPLVYDVTVAAYMSASFSANSQVILTYGQPDVY-DELTYTNS--- 471
 QY 592 DPTCKTEFWATILDSAGPL---PDYTPPEIFSE--STGFTLYGMLYVPHDLOPKKY 646
 DB 472 ----TKPL-RTITDNKVLQIKDVALPNITTYFELPLPSETLANVQRLPRGSPDKKY 525
 QY 647 PVLFTIGGPQVQVUNNRFKGVKK--FRLNTLASIGYVVVINDRGSCHRGKKEGAFKYY 705
 DB 526 PILFTPGGAGAEVTKRQALNFKAVVADSELEYVTWVDNRGTGFKRKRFSAVTRQ 595
 QY 706 MGOIIDDQVEGLQYLASRDFIDLDRVGIHMSYGYVLSLMLMQRSDIFRVALAGAPV 765
 DB 586 LGLLHAEQIYAAQO-AANIPWIDADHIGIMGMSFGIYLSKYLEKDSGAFLLGVITAPV 644
 QY 766 TLMIFDTGTTEYMGHPDNEQGYVIGYVAMQAEKFPSEPNRLLLHGLFDENVHFAAT 825
 DB 645 SDRFSDMSTERYMTLSTNEEGYETSAY-RKTDFKQVVEGFTLLQHGIGDONVHONS 703
 QY 826 SILSLFVBAKGYDQIYPOERHSIRVPSGHEIYH 863

DB 704 AALVDLMDG-----VSPKELHSQWFTDSHGIGISYH 735
 RESULT 3
 US-10-002-593-6
 ; Sequence 6, Application US/10002593
 ; Patent No. 6586198
 ; GENERAL INFORMATION:
 ; APPLICANT: Vanderbilt University
 ; APPLICANT: Brown, Nancy J
 ; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTIT
 ; TITLE OF INVENTION: INHIBITOR AND VASOPEPTIDASE INHIBITOR ASSOCIATED ANGIOEDEMA
 ; FILE REFERENCE: Acty Docker No. 6586198 1242/48/2
 ; CURRENT APPLICATION NUMBER: US/10/002,593
 ; CURRENT FILING DATE: 2001-10-31
 ; PRIOR APPLICATION NUMBER: 60/244,524
 ; PRIOR FILING DATE: 2000-10-31
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6
 ; LENGTH: 766
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-002-593-6
 Query Match 11.3%; Score 529; DB 4; Length 766;
 Best Local Similarity 26.5%; Pred. No. 2,1e-44;
 Matches 186; Conservative 109; Mismatches 280; Indels 128; Gaps 32;
 QY 206 PADPDWI-----AFIHSNDIWSNIYTREERRLLTYVHNELANNEEDARSAGVATFV 256
 DB 149 PNNQWYVWSPVGHKLAYVWNNDIYVKLEPNLPYRITWTG-----KEDITIYNGITDMV 202
 QY 257 LOEE-PDRYSGYWCRAETTPSGKILRLIYEENDESEVEIITH---VTSMLERRADs 312
 DB 203 YEEVFSAYSLAMSPGTF-----LAVAFNDTEVPLIEVSFSDSLQYPKTVR 253
 QY 313 FRYPKGTANPKVTFKXSEIMIDAEGRIIDVIDELIOPFELFEGVYIARAGMTPRG 372
 DB 254 VPYKAGVNPVTFV-FV--VNTDLSSTVNTATSQIARASML-GDHYLCDVYWA---- 306
 QY 373 YAMSILDRSQTLQVILISBELFIYEDDVMERQRLIESVPDSVTPLLIYEETDIWIN 432
 DB 307 -----TOERISLOWL-----RLQNY--SYMIDCVDDESGRW-N 338
 QY 433 IHDIHFVFPQSHHEIEFIASECKTGFRH-----LYKITSLKSKYKSSGGLPAP 485
 DB 339 C-----LVAQHILMSTTGWGRFRPSEPHITLDGNSFYKILS--NEGYRHI----- 384
 QY 486 SDFKCPIKEEIAITSGEMEVILGRHSNIQVDEVRIIVYFEGT-KDSPLEHLLYVVSYPN 544
 DB 385 CYFOIDKDKCTFITKGTWEVIG---IEALTSDYLYIISNEYKMGCGGNLYKIQILIDY 439
 QY 545 GEYTRLDRYSHSCCS--OHCDFPIKSKSNOKNPH--C---VSLYKLSPEDDPTCK 596
 DB 448 TKVYTCI-----SCEINPERCOYVSFSFEKAKYVOLRCSGRGLPYTLTHSSVNDGKL 492
 QY 597 TKEFWATILDSAGPLPDYTPB---IPSESTGTGFTLYGMLYVPHDLOPKKYPTVLY 653
 DB 493 VLED-NSALDKM--LDNVQMPSKLDPITLINEKF--WQVILPRH-FDKSKKYPLLLDVY 547
 QY 654 GGPQVQVUNNRFKGVKFRLN--TLASLGYVVV--INDRGSCHRGKKEGAFKYYKMQI 709
 DB 548 AGPCSQ-----KADTVFRILNWTYLASTENIIVASPDGRGSGYQGGKIHAINRRIGTF 601
 QY 710 EIDDOVEGLQYLASRDFIDLDRVGIHMSYGYVLSLMLMQRSDIFRVALAGAPVTLMT 769
 DB 602 EVADQLEAARQF-SKNGFVNNKRIALIWMSYGYVSMVLGSGSGVFKCGIAVAPSRWE 660
 QY 770 FYDTGYTERYMG--HPDNEQGYVIGYVAMQAEKFPSEPNRLLLHGLFDENVHFAAT 827
 DB 661 YYDSVYTERYMGFLPTEDNLHDYRNSITVMSRAENF--KQVEYILLIGTADDNVHFOCSAO 718

QY 828 LSLFLVAGKPYDLOIYPOERSIRVPESEGEHELHLYLOE 870
 Db 719 ISGALVUVGVDFQAMWTTDEBDHGIASTAHQIYTHMSHFIKQ 761

RESULT 4

PCT-US93-07923-3
 / Sequence 3, Application PC/TUS9307923
 / GENERAL INFORMATION:
 / APPLICANT: Morimoto, Chikao
 / APPLICANT: Schlossman, Stuart F.
 / APPLICANT: Tanaka, Toshiaki
 / TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
 / NUMBER OF SEQUENCES: 16
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Fish & Richardson
 / STREET: 225 Franklin Street
 / CITY: Boston
 / STATE: Massachusetts
 / COUNTRY: U.S.A.
 / ZIP: 02110-2804
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: 3.5" Diskette, 1.44 MB
 / COMPUTER: IBM PS/2 Model 502 or 55SX
 / OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
 / SOFTWARE: Wordperfect (Version 5.0)
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: PCT/US93/07923
 / FILING DATE: 19930819
 / CLASSIFICATION:
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 07/934,162
 / FILING DATE: 21-AUG-1992
 / APPLICATION NUMBER: 07/832,211
 / FILING DATE: 06-FEB-1992
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Fraser, Janis K.
 / REGISTRATION NUMBER: 34,819
 / REFERENCE/DOCKET NUMBER: 00530/055002
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (617) 542-5070
 / TELEFAX: (617) 542-8906
 / TELEX: 200154
 / INFORMATION FOR SEQ ID NO: 3:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 755
 / TYPE: amino acid
 / STRANDEDNESS:
 / TOPOLOGY: linear
 / PCT-US93-07923-3

Query Match 11.2%; Score 525; DB 5; Length 755;
 Best Local Similarity 26.3%; Pred. No. 5.4e-44;
 Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;
 QY 206 PADPWI-----AFHSNDIMVSNITVREBRRLTVNHEMLNMEPDASAGVATVY 256
 Db 138 PNTQWVWVMSPGHKLAVYWNNDIYKIEPNIPSTRITWG-----KEDITYNGITDWV 191
 QY 257 LOEB-FDRYSGYWCPCAKETTPSGGKILNIIYEENDESEVEIHH--VTSPLMIFRRADS 312
 Db 192 YEEVEFSAVSALMWSPNGTF-----LAVAGNDTEVPLIEVSFYSDESLQYFKYR 242
 QY 313 FRYEPTGTANPKVTFRKSEIMIDAGRIIDVDKELIOFELLFGVEYIAPAGWTPBECK 372
 Db 243 VYEPAGAVNPFVKFV--VNTDSLSSVTNATSIQITAPASMLI-GDHYLCDVTA--- 295
 QY 373 YAMSIILDRSQTRLOIVLISPELFIPEVDVWERQRLIESVDSVTPLIYETTDIMIN 432
 Db 296 -----TQERISLOWL-----RRIQNY--SYMDICDDESSGRW-N 327
 QY 433 IHDIFHVPQSHHEEIEPIFASECKTGRH-----LYKITSIKESKYSKSSGGLPAP 485

Db 328 C-----LVANQHIEEMSTGWVGRFRPSEPHFTLDGNSFYKII--NEEGYRHI----- 373
 QY 486 SDFKCPYKEEIAITSGEMEYLGHRGSMIODEVARLYFPEGT-KDSPLHEHLYVVSYPN 544
 Db 374 CYFOIDKKDCFTFIKGTWEVIG-----IEALTSIYLYISNEYKMGFGGRNLYKI----- 423
 QY 545 GEVTRLTDRGYSHSCCIS-----QHCDFFISKYSNQKPH---C-----VSLYKLSPEDD 592
 Db 424 ----QLSD--YTKVYTLSCSELNPERCQYVSFSKSKAKYQLRSGSGGLPYTLHSSVND 477
 QY 593 PCTKTEFMATIIIDSGAPLDIYPP--IFSESTGFTLYGMLYKPHDLOPKKPYTV 649
 Db 478 KGLRVLED-NSALDKM--LQNVQMPSEKDLFIILNETKF-WYQMIILPQH-FDKSKKPYL 532
 QY 650 LFTYGGPOVLANNRFGVAYFRLLN---TLASLGVVVV-IDNNGSGHRGLKFGAPKX 705
 Db 533 LDVYAGPCSQ-----KADTVFRUNWATYLASTENITVASPDGSGSYGDKIMHAINR 586
 QY 706 MGQIEIDQVEGLQYLASRYDFIDLVRGIGHGVSYGGLSLMALMQRSDIFRVAIAGAPV 765
 Db 587 LGTFEVEDQIEARQF-SKMGFVDNKRIALMGWSYGYVSMVLGSSGSGVFKGIAVAPV 645
 QY 766 TLMIFYDTGTERYMG--HPDQEGGYLGASVAMQEKFPSEPNRLILHGFLENVHFA 823
 Db 646 SRMEYDVSYTERYMGLPPEPEDI,LDHYRNSTVMSRAENF--KQVEYLLIHGTADNVHFQ 703
 QY 824 HTSILSLFVAGKPYDLOIYPOERSIRVPESEGEHELHLYLOE 870
 Db 704 QSAQISKALVGVDFQAMWTTDEBDHGIASTAHQIYTHMSHFIKQ 750

RESULT 5

PCT-US93-07923-2
 / Sequence 2, Application PC/TUS9307923
 / GENERAL INFORMATION:
 / APPLICANT: Morimoto, Chikao
 / APPLICANT: Schlossman, Stuart F.
 / APPLICANT: Tanaka, Toshiaki
 / TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
 / NUMBER OF SEQUENCES: 16
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Fish & Richardson
 / STREET: 225 Franklin Street
 / CITY: Boston
 / STATE: Massachusetts
 / COUNTRY: U.S.A.
 / ZIP: 02110-2804
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: 3.5" Diskette, 1.44 MB
 / COMPUTER: IBM PS/2 Model 502 or 55SX
 / OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
 / SOFTWARE: Wordperfect (Version 5.0)
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: PCT/US93/07923
 / FILING DATE: 19930819
 / CLASSIFICATION:
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 07/934,162
 / FILING DATE: 21-AUG-1992
 / APPLICATION NUMBER: 07/832,211
 / FILING DATE: 06-FEB-1992
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Fraser, Janis K.
 / REGISTRATION NUMBER: 34,819
 / REFERENCE/DOCKET NUMBER: 00530/055002
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (617) 542-5070
 / TELEFAX: (617) 542-8906
 / TELEX: 200154
 / INFORMATION FOR SEQ ID NO: 2:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 759
 / TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07923-2

Query Match 11.2%; Score 525; DB 5; Length 759;
Best Local Similarity 26.3%; Pred No. 5.4e-44;
Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

QY 206 PADPDMT-----AFHSNDIWSNIVTREERLTYVHNLANMEDARSAGVATFV 256
DB 142 FNTQWMTWSPVGHKLAVYMNNDIYKLEPNLPSRITWTG-----KEDIYNGITDWMV 195
QY 257 LQEF-FDRYSGYWCPCAKETTPSGKILRILEYNDESEVEIHH---VTSPLLETRADS 312
DB 196 YEEVFASYSALWMSPNGT-----LAAQFNDEVPDLIEYSFSDLSLOPKTVR 246
QY 313 FRYPTGTANPKVTFKMSIMIDAGRLLIDVLDKLIQPELLEGEVYIARAGTPEBK 372
DB 247 VPIPKAGVNPVTFV--VNTDSLSSVTNATSIQITPASMIL-GDHYLCDVTWA--- 299
QY 373 YAMSLIDRSQTRLQIVLISPELPIVEDDWERORLIESVPDSVTPLIYEETDWIN 432
DB 300 -----TOERISLQWL-----RRIQNV--SVMDICDDESSGR-N 331
QY 433 IHDFHVPQSHHEEIEFIFASECKTGFRH-----LYKITSILKSKYKSSGGLPAP 485
DB 332 C-----LVARQHLEMTSTGWGRFPSEPHFTLDGNSFYKIIIS--NEGYRHI----- 377
QY 486 SDFKCPKEIEIATSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHLLVYVSYNP 544
DB 378 CYFQIDKKDCFTIKGTWEVIG-----LEALTSQVLYIISNEYKMGNGRNLTKI----- 427
QY 545 GEVTRLDRGYSHSCCIS-----QHCDFFISKYSNOKNPH---C-----VSLYLKLSPEDD 592
DB 428 ----QLSD--YTRVYTCISCEINPERCQYYSVSFSKAKYQOLRCSGRLPLTYLTHSSVND 481
QY 593 PTCKTEKFWATILDSAGRLPDYTPPE---IFSESTGTFTLYKMLYKPHDLQPKKPTV 649
DB 482 KGLRVLEB-NSALDKM--LQNVQMPSKLDFTILNETKF-WYQMLPPH-PDKSKKPYLL 536
QY 650 LFIYGGPOVOLVNNRFKGVKYFRLLN--TLASLGVVVV--IDNRGSGRGLKFGAGAKYK 705
DB 537 LDVYAGPCSQ-----KADTVFRLNMAITYLASTENIIVASFDGSGSGQGGKIMEALNRR 590
QY 706 MGQIEIDQVBEGLQYLASRYDFIDLDRVGIHGSYGYSLMALMQRSDIFRVAIAGAPV 765
DB 591 LGTFEVEDQIEARQF-SKMGFVDNKRRIAIWMSYGYVTSVILGSSGSGVFKCGIAVAPV 649
QY 766 TLMIFYDTGTYTERYMG--HPDQNEGGYVIGSVAMQAEKPESENNRLLHLHGFIDENYHFA 823
DB 650 SRWEYDSVYTERKMGFLTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 707
QY 824 HTSILSLFLVAKRPYDLQIYPOERHSIRVSEGEHYELHLHLYLOE 870
DB 708 QSAQISKALVDVGDVQAMWYTDDEHGLASSTAHQHLYTHMSHIKQ 754

RESULT 6

US-08-230-491A-3
Sequence 3, Application us/08230491A
Patent No. 5587299

GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
TITLE OF INVENTION: ISOLATED NOCLETIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN, AND USES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: FELPE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK

COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: us/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: IUD 330
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-230-491A-3

Query Match 11.2%; Score 525; DB 1; Length 766;
Best Local Similarity 26.3%; Pred No. 5.5e-44;
Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

QY 206 PADPDMT-----AFHSNDIWSNIVTREERLTYVHNLANMEDARSAGVATFV 256
DB 149 FNTQWMTWSPVGHKLAVYMNNDIYKLEPNLPSRITWTG-----KEDIYNGITDWMV 202
QY 257 LQEF-FDRYSGYWCPCAKETTPSGKILRILEYNDESEVEIHH---VTSPLLETRADS 312
DB 203 YEEVFASYSALWMSPNGT-----LAAQFNDEVPDLIEYSFSDLSLOPKTVR 253
QY 313 FRYPTGTANPKVTFKMSIMIDAGRLLIDVLDKLIQPELLEGEVYIARAGTPEBK 372
DB 254 VPIPKAGVNPVTFV--VNTDSLSSVTNATSIQITPASMIL-GDHYLCDVTWA--- 306
QY 373 YAMSLIDRSQTRLQIVLISPELPIVEDDWERORLIESVPDSVTPLIYEETDWIN 432
DB 307 -----TOERISLQWL-----RRIQNV--SVMDICDDESSGR-N 338
QY 433 IHDFHVPQSHHEEIEFIFASECKTGFRH-----LYKITSILKSKYKSSGGLPAP 485
DB 339 C-----LVARQHLEMTSTGWGRFPSEPHFTLDGNSFYKIIIS--NEGYRHI----- 384
QY 486 SDFKCPKEIEIATSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHLLVYVSYNP 544
DB 385 CYFQIDKKDCFTIKGTWEVIG-----LEALTSQVLYIISNEYKMGNGRNLTKI----- 434
QY 545 GEVTRLDRGYSHSCCIS-----QHCDFFISKYSNOKNPH---C-----VSLYLKLSPEDD 592
DB 435 ----QLSD--YTRVYTCISCEINPERCQYYSVSFSKAKYQOLRCSGRLPLTYLTHSSVND 488
QY 593 PTCKTEKFWATILDSAGRLPDYTPPE---IFSESTGTFTLYKMLYKPHDLQPKKPTV 649
DB 489 KGLRVLEB-NSALDKM--LQNVQMPSKLDFTILNETKF-WYQMLPPH-PDKSKKPYLL 543
QY 650 LFIYGGPOVOLVNNRFKGVKYFRLLN--TLASLGVVVV--IDNRGSGRGLKFGAGAKYK 705
DB 544 LDVYAGPCSQ-----KADTVFRLNMAITYLASTENIIVASFDGSGSGQGGKIMEALNRR 597
QY 706 MGQIEIDQVBEGLQYLASRYDFIDLDRVGIHGSYGYSLMALMQRSDIFRVAIAGAPV 765
DB 598 LGTFEVEDQIEARQF-SKMGFVDNKRRIAIWMSYGYVTSVILGSSGSGVFKCGIAVAPV 656
QY 766 TLMIFYDTGTYTERYMG--HPDQNEGGYVIGSVAMQAEKPESENNRLLHLHGFIDENYHFA 823
DB 657 SRWEYDSVYTERKMGFLTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 714

QY 824 HTSILSLFLVRACKPYDLOIYPOERSIRVPESEGEHELHLYLOE 870
 Db 715 QSAQISKALVDVGVDFOAMWYTDHGHIASTAHQIYTHMSHFIKQ 761

RESULT 7

US-08-619-280A-3
 ; Sequence 3, Application US/08619280A
 ; Patent No. 5767242
 ; GENERAL INFORMATION:
 ; APPLICANT: Zimmermann, Rainer; Park, John E.;
 ; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
 ; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
 ; TITLE OF INVENTION: ALPHA, AND USES THEREOF
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felte & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/619,280A
 ; FILING DATE: 18-MARCH-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/230,491
 ; FILING DATE: 20-APRIL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 5767242man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5330.1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 838-3884
 ; TELEFAX: (212) 838-9200
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 766 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-619-280A-3

Query Match 11.2% Score 525; DB 1; Length 766;
 Best Local Similarity 26.3%; Pred. No. 5.5e-44;
 Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

QY 206 PADPDMW-----AFHSNDIWNISNIVTREERLTYVHNELANMEDARSAGVATFV 256
 Db 149 PNNQWVWSPVGHKLAVYNNNDIYVKLEPNLPSYRITWTG-----KEDIYNGIDWV 202
 QY 257 LQEE-FDRYSGYWMCPKAEPTTPSGKILRLIYEENDESEVHIH---VTSMLETRADS 312
 Db 203 YEEVFSAISALWSPNCTF-----LVAQNDTEVPLIEVSFSDSLDYPKTVR 253
 QY 313 FRYPKTGTPANPKVTPKMSIMIDAGRIIDVIDKLIOPFILFEGVEYIARAQMTPEGK 372
 Db 254 VYPRAGAVNTPVKFFV--VNTDSLSTVNTATSIQITAPAMLI-GDHYLCDVYMA--- 306
 QY 373 YAMSLIDRSQTRLOIVLISBELFLPVEDDVMERQRLIESVPDSVTPILIIYEETDIWIN 432
 Db 307 -----TQERISLQWL-----RQYNY--SYMDICVDESSEGRW-N 338
 QY 433 IHDIFVFPQSHHEEIEIFASECTGPRH-----LYKTSILKESKYRSRSGGAP 485
 Db 339 C-----LVARQHIEMSTTGWVGRFRPSSEPHFLDOGNSFYKLIIS--NEEGYHIL----- 384
 QY 486 SDFKCPIMEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHLLVYVSYNP 544

Db 385 CFQIDKKDCTPTTKGTWEIG-----TEALSDIYYISNEYKMPGGRNLTKI----- 434
 QY 545 GEVYRLDRGYSKSCIS-----QHCFPIRSKSNQKNP---C-----VSLYKLSPEDD 592
 Db 435 ----QLSD--YTKVCTLSCELNERCQYVSFSKAKYQJLRCSGPGPLTYLHSSVND 488
 QY 593 PCKTKEFMATILDSAGPLPDYTPRE---IFSESTGTGLYGMLYKPHDLQPKKYPV 649
 Db 489 KGLRVLED-NSALDKM--LQNVQMPSKLDPIILNETKT-WYQMLIPH-PDKSKYPLL 543
 QY 650 LPIYGPQVQVLVNNRFGYKFFRLN--TLASIGYVVV-IDNRGSGHRLKEGAFKKY 705
 Db 544 LDVYAGPCG-----KADTVRELMATYLAISENITVASFDRSGSGYGDGKIMAINR 597
 QY 706 MGQIEIDQVEGLQYASRYDPIIDRVGIGHWSYGYSLMALMQRSDIFRYAIAQAV 765
 Db 598 LGTFEEVDQTEARQF-SKQGFVDNRIRAIWGSYGYATSWLGSQGVFKGIAVAVP 656
 QY 766 TLMIFPDGTGYTERYMG--HPDQNEGYLYGSYVMAQEKFPSEBNRLLLHGFLENVHA 823
 Db 657 SRMEYDQVYTERYMGELPTPEDNLDHYRNSVMSRAENF--KQEVLLIHGTADNVHQQ 714
 QY 824 HTSILSLFLVRACKPYDLOIYPOERSIRVPESEGEHELHLYLOE 870
 Db 715 QSAQISKALVDVGVDFOAMWYTDHGHIASTAHQIYTHMSHFIKQ 761

RESULT 8

US-08-940-391-3
 ; Sequence 3, Application US/08940391
 ; Patent No. 5965373
 ; GENERAL INFORMATION:
 ; APPLICANT: Zimmermann, Rainer; Park, John E.;
 ; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
 ; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
 ; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felte & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/940,391
 ; FILING DATE: 01-OCT-1997
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/619,280
 ; FILING DATE: 18-MARCH-1996
 ; APPLICATION NUMBER: 08/230,491
 ; FILING DATE: 20-APRIL-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 5965373man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5330.1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 838-3884
 ; TELEFAX: (212) 838-9200
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 766 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-940-391-3

Query Match 11.2%; Score 525; DB 2; Length 766;
 Best Local Similarity 26.3%; Pred. No. 5, 5e-44;
 Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

```

QY 206 PADPDMT-----AFHSNDIWSNVITREERLITVYHNLAMMEDASAGVATPV 256
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 PNNQWVTVSPVGHKLAYWNNDIYKIEPNLPSYRITWTG-----KEDIYNGITDVI 202
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 LQEE-FDRYSGWMCCKAETTPSGKILRLYEENDESEVEIHH--VTSPLMLETTRADS 312
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 YEEVFSAYSALWMSPNGT-----LAVAQFNDTEVPLIEVSFYDESIOYKQTVR 253
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 FRYPKGTANPKVTEKMESEIMIDAEGRIIDVDKELIOPPELLFEGVEYIARAGMTPECK 372
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 VYPRKAGAVNPVYKFFV--VNTDSLSVTNATSIQTAPASMLI--GDHYLCVYTA----- 306
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 373 YAMSILDRSQRLQIVLISPELLFIVEDDVMERQRLIESVDSVTPPLIYEETDIDWIN 432
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 -----TQERISLQWL-----RRIQNY--SVWIDICDYDESSGRW-N 338
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 IHDIHVFPQSHHEEIEFIASECKTGFRH-----LYKTSILKESKYKSSGGLPAP 485
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 C-----LVARQHIEMSTTGWGRFPPSEPHFTLDGNSFYKIIIS--NEEGYRHI----- 384
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 SDFKCPIKEELAITSGEMEVLRHGSNIQVDEVRRLVFEET-KDSPLHHLVYVSYNVP 544
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 CYFOIDKDKDCTFTTGTWEVIG-----IEALTSVLYIYSNEYKMGMPGGRNLYKI----- 434
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 545 GEVTRLTRGYSHSCCIS-----QHDFEISKYSNQNKNPH--C-----VSLYKLSPPEDD 592
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 ----QLSD--YTKVCLSCELNPERCQYVSFSKAKYQJLRGSGPLPLTYTLHSSVND 488
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 593 PTCKTEKFWATILDSAGPLPDYTPPE--ISFESTGTFTLYGMLYKHHDLOPKKYPV 649
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 KGLRFLBD-NSALDQM--LQNVQMPSSKLDPIILNETKF-WYQMLLPFH-FDKSKYPL 543
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 650 LFIYGGPOVOLVNNRFKGVKYFRLN--TLASLGYVVVV-IDNRGSCRGKLFEGAFYKX 705
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 544 LDVYAGPSCQ-----KADYVFLNMAIYASTENIIVASFDGRSGYQGDKIMAIINRR 597
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 706 MGQLEIDQVGLQYLASRYDFIDLRVGIHGSYGYLSMALMQRSDFRVAIAGAPV 765
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 598 LGTEFEVDQIEAARQF-SKMGFVDNKRIAIWMSYGVYTSVGLSGSGVFCGIAVAPV 656
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 766 TLMIFYDTGYTERYMG--HPDQNEGYLYGSVAMQAEKFPSEPNRLILHGFLENVHFA 823
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 657 SRMEYISVYTERIMGLPTPEDNLDHYRNSITWMSRAENF--KQVEYLLIHGTADDNVHFQ 714
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 824 HTSILSLFLVAGKPYDLQIYPOERHSIRVESGHEYLHLHYLOE 870
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 QSAQISKALVDVGVDFQAMWYTTDEDHGLASTAHQHITYHMSHPKIQ 761
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9
 US-09-794-236-1
 ; Sequence 1, Application US/09794236
 ; Patent No. 6337069
 ; GENERAL INFORMATION:

; APPLICANT: Grouzmann, Eric
 ; APPLICANT: Lacroix, Jean-Silvain
 ; APPLICANT: Monod, Michel
 ; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
 ; FILE REFERENCE: 81985/276823
 ; CURRENT APPLICATION NUMBER: US/09/794.236
 ; CURRENT FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 766
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-794-236-1

Query Match 11.2%; Score 525; DB 4; Length 766;
 Best Local Similarity 26.3%; Pred. No. 5, 5e-44;
 Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

```

QY 206 PADPDMT-----AFHSNDIWSNVITREERLITVYHNLAMMEDASAGVATPV 256
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 PNNQWVTVSPVGHKLAYWNNDIYKIEPNLPSYRITWTG-----KEDIYNGITDVI 202
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 LQEE-FDRYSGWMCCKAETTPSGKILRLYEENDESEVEIHH--VTSPLMLETTRADS 312
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 YEEVFSAYSALWMSPNGT-----LAVAQFNDTEVPLIEVSFYDESIOYKQTVR 253
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 FRYPKGTANPKVTEKMESEIMIDAEGRIIDVDKELIOPPELLFEGVEYIARAGMTPECK 372
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 VYPRKAGAVNPVYKFFV--VNTDSLSVTNATSIQTAPASMLI--GDHYLCVYTA----- 306
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 373 YAMSILDRSQRLQIVLISPELLFIVEDDVMERQRLIESVDSVTPPLIYEETDIDWIN 432
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 -----TQERISLQWL-----RRIQNY--SVWIDICDYDESSGRW-N 338
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 IHDIHVFPQSHHEEIEFIASECKTGFRH-----LYKTSILKESKYKSSGGLPAP 485
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 C-----LVARQHIEMSTTGWGRFPPSEPHFTLDGNSFYKIIIS--NEEGYRHI----- 384
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 SDFKCPIKEELAITSGEMEVLRHGSNIQVDEVRRLVFEET-KDSPLHHLVYVSYNVP 544
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 CYFOIDKDKDCTFTTGTWEVIG-----IEALTSVLYIYSNEYKMGMPGGRNLYKI----- 434
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 545 GEVTRLTRGYSHSCCIS-----QHDFEISKYSNQNKNPH--C-----VSLYKLSPPEDD 592
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 ----QLSD--YTKVCLSCELNPERCQYVSFSKAKYQJLRGSGPLPLTYTLHSSVND 488
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 593 PTCKTEKFWATILDSAGPLPDYTPPE--ISFESTGTFTLYGMLYKHHDLOPKKYPV 649
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 KGLRFLBD-NSALDQM--LQNVQMPSSKLDPIILNETKF-WYQMLLPFH-FDKSKYPL 543
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 650 LFIYGGPOVOLVNNRFKGVKYFRLN--TLASLGYVVVV-IDNRGSCRGKLFEGAFYKX 705
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 544 LDVYAGPSCQ-----KADYVFLNMAIYASTENIIVASFDGRSGYQGDKIMAIINRR 597
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 706 MGQLEIDQVGLQYLASRYDFIDLRVGIHGSYGYLSMALMQRSDFRVAIAGAPV 765
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 598 LGTEFEVDQIEAARQF-SKMGFVDNKRIAIWMSYGVYTSVGLSGSGVFCGIAVAPV 656
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 766 TLMIFYDTGYTERYMG--HPDQNEGYLYGSVAMQAEKFPSEPNRLILHGFLENVHFA 823
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 657 SRMEYISVYTERIMGLPTPEDNLDHYRNSITWMSRAENF--KQVEYLLIHGTADDNVHFQ 714
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 824 HTSILSLFLVAGKPYDLQIYPOERHSIRVESGHEYLHLHYLOE 870
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 QSAQISKALVDVGVDFQAMWYTTDEDHGLASTAHQHITYHMSHPKIQ 761
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10
 US-08-230-491A-2
 ; Sequence 2, Application US/08230491A
 ; Patent No. 5587299
 ; GENERAL INFORMATION:

; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
 ; APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
 ; TITLE OF INVENTION: ISOLATED NOCTEIC ACID MOLECULE CODING FOR
 ; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN ` AND USES
 ; TITLE OF INVENTION: THEREOF
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: FELFE & LYNCH
 ; STREET: 805 THIRD AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE

```

COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230.491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 688-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-230-491A-2

```

```

Query Match          9.7%; Score 454; DB 1; Length 760;
Best Local Similarity 24.1%; Pred. No. 9.5e-37;
Matches 169; Conservative 113; Mismatches 269; Indels 150; Gaps 27;

```

```

QY 212 IAFHSNDIWNITVREERRLTYVHNELANMEDASAGVATFVLOEEF--DRYSGYWM 269
DB 162 LAYVYONNITVYKORGPDPFOITF-----NGRENKIFNGIPDWYEEMLPTKYA-LWM 214
CY 270 CPKAEITPSGGKILRLIYEENDESEVEIIVHTSPMLET--RRADSFRYPKGTANPKYTK 328
DB 215 SP-----NKKFL--AAAEFNDKDIPVAVSYGDEQYPTTINIPYKAGAKNVVARI- 264
QY 329 MSEIMDAEGRILIVIDKELIOPPEI-----LFEGVEYIARAAGTPEKGYAMSILLDSQ 383
DB 265 -----FLIDTTPAYVGPQEVVPVAMIASDYYFSLTWVTERVCLQWL----- 309
QY 384 TRLQIVLISPELFIPEVDVMERQRLIESVPSVT-----PLIYEETDIIWINI 433
DB 310 KRVONVSVLSICDFREDQMTWDCPKTOEHIESRTGAGGFVPSVPSYDA-----ISY 364
QY 434 HDIFHVFPQSHHEEIEFIASECKTGPHLYKITSILKESYKRSQGLPAPSPKCPK 493
DB 365 YKIF-----SDKDGXKIHVI-----KDTVE 385
QY 494 EELAITSGEWEVLGRHGSNIQVDEVRRLVFEGTKDSPLEHLLVSVYNPGEVTRLTDR 553
DB 386 NAIQITSGKWEAI-----NI-----FRVTDLSLFYSNFEFEYPRGRNIYRISIG 430
QY 554 GYSHS-CCISQH-----CDFISKYSNQNPHCVSLY-----KLSPEDDDPTCKTEPMAT 603
DB 431 SYPSKACVTCFLAKKRCQYTTASFSDAKYVALVCYGPGLPISTLHDGHTDQ---IK 486
QY 604 ILDSAGLPD-----YTPPEIFSFESTTGFLLYGMLYKPPHDLQPKYPTVLFYGGPQV 658
DB 487 ILLENKELEMLAKNIQJPKKEIKLEVDETILWYKMLLPQPDRSKXYPLLIQYGGPQS 546
QY 659 QLVNRRFPGVAYFFLNTLASL-----GYVVVYINDRGCHGKLEKRGFKYKMGCIHIDDQ 714
DB 547 QSVSS-----VFANWMLSYLASKRGWIALVDRGTAFQDDKLLVAVYRKLGYVEEDQ 600
QY 715 VEGQYLASRYDFIDLDRVGIHGWSYGYSLMALMORSDFRVAIAGAVPTLMIFFYDTG 774
DB 601 ITAVRKFI-EMGFIDEXRIALIMGWSYGYVSLALAGSTGLFKGGLAVANVSSMEYVASY 659
QY 775 YTERYMGHP--DQNEQGYILGSVAMQAKFPSENNRLLILGLFDEVNFAHTSILISFL 832
DB 660 YTERFMGLPTDQDLLEHKYKSTVWARAEYFRNVD--YLLIHGTADNVHFONSQIAKAL 717
QY 833 VRAGKPYDLQIYPERHSIRVPSG---EHYELHLLAYLOE 870
DB 718 VNAQVDPQAMWYSDQNHGL-----SGLSTNHLTYTHMTHFLKQ 754

```

RESULT 11

```

US-08-619-280A-2
Sequence 2, Application US/08619280A
Patent No. 5767242
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
APPLICANT: Retig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
TITLE OF INVENTION: ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSER: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,280A
FILING DATE: 18-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5767242man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-619-280A-2

```

```

Query Match          9.7%; Score 454; DB 1; Length 760;
Best Local Similarity 24.1%; Pred. No. 9.5e-37;
Matches 169; Conservative 113; Mismatches 269; Indels 150; Gaps 27;

```

```

QY 212 IAFHSNDIWNITVREERRLTYVHNELANMEDASAGVATFVLOEEF--DRYSGYWM 269
DB 162 LAYVYONNITVYKORGPDPFOITF-----NGRENKIFNGIPDWYEEMLPTKYA-LWM 214
CY 270 CPKAEITPSGGKILRLIYEENDESEVEIIVHTSPMLET--RRADSFRYPKGTANPKYTK 328
DB 215 SP-----NKKFL--AAAEFNDKDIPVAVSYGDEQYPTTINIPYKAGAKNVVARI- 264
QY 329 MSEIMDAEGRILIVIDKELIOPPEI-----LFEGVEYIARAAGTPEKGYAMSILLDSQ 383
DB 265 -----FLIDTTPAYVGPQEVVPVAMIASDYYFSLTWVTERVCLQWL----- 309
QY 384 TRLQIVLISPELFIPEVDVMERQRLIESVPSVT-----PLIYEETDIIWINI 433
DB 310 KRVONVSVLSICDFREDQMTWDCPKTOEHIESRTGAGGFVPSVPSYDA-----ISY 364
QY 434 HDIFHVFPQSHHEEIEFIASECKTGPHLYKITSILKESYKRSQGLPAPSPKCPK 493
DB 365 YKIF-----SDKDGXKIHVI-----KDTVE 385
QY 494 EELAITSGEWEVLGRHGSNIQVDEVRRLVFEGTKDSPLEHLLVSVYNPGEVTRLTDR 553
DB 386 NAIQITSGKWEAI-----NI-----FRVTDLSLFYSNFEFEYPRGRNIYRISIG 430
QY 554 GYSHS-CCISQH-----CDFISKYSNQNPHCVSLY-----KLSPEDDDPTCKTEPMAT 603

```



```

Db 411 SYPSKCCVCHLKRERCOYTTASSDYAKYALVCYGPPISTLHDKRTDQ-----IK 486
Qy 604 ILDSAGLPD-----YTPPEIFSESTGTGFTLYGMLYKPHDLOPKKYPVLYFYGGPQV 658
Db 487 ILEENKEJENMLKNIOQPKKEIKLEVEDEITLWYGMILPPQFDRSKKYPPLIQYGGPCS 546
Qy 659 QLVNNRFGVKYFRINTLASL---GYVVVVIDNRGSGHGLKEFGAFKYMGOIEIDDQ 714
Db 547 QSVRS-----VFAVNMISTYLASKEGWIALVDGRTAFQGDKLLYAVRYRLGYVEVEDQ 600
Qy 715 VEGLOYLASRYDFIDLDVGIHGMVSGYSLMALMQRSDFRVAIAGAPVLTMTFYDTG 774
Db 601 ITAVRKFI-EMGFIDEKRIALWMSYGYVSSIALASGTGLFKCGIAVAPVSSMEYASV 659
Qy 775 YTERYMGHP--DQNEQGYVIGSVAMQAEKPESEPNRLLHGFLENVHFAHTSILSFL 832
Db 660 YTERFMGLPTKDNLEHKNSTVWARAEYFRNVD--YLLHGTADNVHFOQSAQIAKAL 717
Qy 833 VRAGKPYDLOIYPOERHSIRVPESSG---EHYELHLHYLOE 870
Db 718 VNAQVDFQAMWYSQDNHGL---SGLSTNHLTYHMTHTFLKQ 754

```

RESULT 12

```

US-08-940-391-2
; Sequence 2, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.;
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5965373man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-940-391-2

```

```

Query Match      9.7%; Score 454; DB 2; Length 760;
Best Local Similarity 24.1%; Freq. No. 9.5e-37;
Matches 169; Conservative 113; Mismatches 269; Indels 150; Gaps 27;

```

```

Qy 212 IAFHSNDIWNISNIVTBERRLTYVHNELANMEEDASAGVATFYLOE--DRYSGYW 269
Db 162 LAYVQNNIYIKQRPGRPPFOITF-----NGRENKIPNGIPDWYEEEMPTKYA-LMW 214
Qy 270 CPKAEPTPSGKILRIIYEENDESEVELIHVTSFMLET-RRASFRYKPTGTANPKVTFK 328
Db 215 SP-----NGKFL--AAEFNDKIDPIVAVSYGDEQYPTINIPYKAAKNPVNRI- 264
Qy 329 MSEIMDAEGIIIVIDKELIQPEI-----LFEQVEYIAPAGVTPSGKYMMSILDRSQ 383
Db 265 -----FIDITTPAYVPOEVPVPMIASSDYFSLWTWTERVCLQWL----- 309
Qy 384 TRLOIVLISPLRFIVEDDWERQRLTESVDSVT-----PLIYETTDIWNINI 433
Db 310 KRVQNVSLASICDREDQWTDCKTQDHEIEKSTGMAGFVSRRVSYDA-----ISY 364
Qy 434 HDIFHVPQSHHEEIEFIASECKTGRHLKYKITSILKESKYSKSSGGLPAPSPKCPK 493
Db 365 YKIF-----SDKGYKHIMYI-----KDTVE 385
Qy 494 EELIATSGEWEVLGRHSNIOVDEVRRLVFEQTKDSPLEHMLVYVSYVNPGEVTRLTDR 553
Db 386 NAIQITSGKWEAL-----NT-----FRVQDSLFFSSNFEPEYGRRNIRISIG 430
Qy 554 GYSHS-CCISQH-----CDFEISKYSNQKXPHCVSLY-----KLSSPDDPTCKTKEMAT 603
Db 431 SYPSKCCVCHLKRERCOYTTASFSYAKYALVCYGPPISTLHDKRTDQ-----IK 486
Qy 604 ILDSAGLPD-----YTPPEIFSESTGTGFTLYGMLYKPHDLOPKKYPVLYFYGGPQV 658
Db 487 ILEENKEJENMLKNIOQPKKEIKLEVEDEITLWYGMILPPQFDRSKKYPPLIQYGGPCS 546
Qy 659 QLVNNRFGVKYFRINTLASL---GYVVVVIDNRGSGHGLKEFGAFKYMGOIEIDDQ 714
Db 547 QSVRS-----VFAVNMISTYLASKEGWIALVDGRTAFQGDKLLYAVRYRLGYVEVEDQ 600
Qy 715 VEGLOYLASRYDFIDLDVGIHGMVSGYSLMALMQRSDFRVAIAGAPVLTMTFYDTG 774
Db 601 ITAVRKFI-EMGFIDEKRIALWMSYGYVSSIALASGTGLFKCGIAVAPVSSMEYASV 659
Qy 775 YTERYMGHP--DQNEQGYVIGSVAMQAEKPESEPNRLLHGFLENVHFAHTSILSFL 832
Db 660 YTERFMGLPTKDNLEHKNSTVWARAEYFRNVD--YLLHGTADNVHFOQSAQIAKAL 717
Qy 833 VRAGKPYDLOIYPOERHSIRVPESSG---EHYELHLHYLOE 870
Db 718 VNAQVDFQAMWYSQDNHGL---SGLSTNHLTYHMTHTFLKQ 754

```

RESULT 13

```

US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Kazuhiko
; TITLE OF INVENTION: THEMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; EARLIER FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-016-080-1

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:13:57 ; Search time 124.626 Seconds
(without alignments)
1140.337 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700
Sequence: 1 MAAMETEQGVETFTADG.....HLHYQENLGSRIALKVI 882

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4700	100.0	882	10	US-09-976-674-1 Sequence 1, Appl
2	4700	100.0	882	12	US-10-054-776-2 Sequence 2, Appl
3	4700	100.0	882	12	US-10-170-789-38 Sequence 38, Appl
4	3504	74.6	658	10	US-09-976-674-19 Sequence 19, Appl
5	3504	74.6	661	10	US-09-976-674-11 Sequence 11, Appl
6	3504	74.6	690	10	US-09-976-674-71 Sequence 71, Appl
7	3236	68.9	613	10	US-09-976-674-21 Sequence 21, Appl
8	2870	61.1	863	10	US-09-976-674-3 Sequence 3, Appl
9	2870	61.1	892	10	US-09-976-674-23 Sequence 23, Appl
10	2870	61.1	892	10	US-09-976-674-27 Sequence 27, Appl
11	2820.5	60.0	879	10	US-09-976-674-35 Sequence 35, Appl
12	2820.5	60.0	879	10	US-09-976-674-33 Sequence 33, Appl
13	2406	51.2	832	10	US-09-976-674-29 Sequence 29, Appl
14	2406	51.2	832	10	US-09-976-674-31 Sequence 31, Appl
15	2356.5	50.1	819	10	US-09-976-674-37 Sequence 37, Appl

16	2356.5	50.1	819	10	US-09-976-674-39 Sequence 39, Appl
17	1808	38.5	358	10	US-09-976-674-13 Sequence 13, Appl
18	1645.5	35.0	310	12	US-09-993-959-4 Sequence 4, Appl
19	1643.5	35.0	508	15	US-10-237-271-3 Sequence 4, Appl
20	1391	29.6	518	10	US-09-976-674-25 Sequence 25, Appl
21	1278	27.2	241	10	US-09-976-674-9 Sequence 9, Appl
22	1049	22.3	432	15	US-10-237-271-4 Sequence 9, Appl
23	1007.5	21.4	194	10	US-09-976-674-17 Sequence 17, Appl
24	585	12.4	129	11	US-09-764-891-3564 Sequence 3564, Ap
25	529	11.3	766	12	US-10-423-714-6 Sequence 6, Appl
26	529	11.3	766	14	US-10-002-593-6 Sequence 6, Appl
27	529	11.3	766	15	US-10-165-603-7 Sequence 7, Appl
28	525	11.2	766	9	US-09-265-608-3 Sequence 3, Appl
29	525	11.2	766	12	US-09-993-959-1 Sequence 1, Appl
30	523.5	11.1	767	15	US-10-165-603-4 Sequence 4, Appl
31	495	10.5	710	15	US-10-156-761-10681 Sequence 10681, A
32	484.5	10.3	818	12	US-10-401-437-3 Sequence 3, Appl
33	484.5	10.3	818	12	US-10-402-067-3 Sequence 3, Appl
34	484.5	10.3	818	12	US-10-401-436-3 Sequence 3, Appl
35	484.5	10.3	818	12	US-09-976-674-15 Sequence 15, Appl
36	464	9.9	108	10	US-09-976-674-41 Sequence 41, Appl
37	461	9.8	706	10	US-09-976-674-5 Sequence 5, Appl
38	461	9.8	796	10	US-09-976-674-5 Sequence 5, Appl
39	461	9.7	760	12	US-10-301-822-55 Sequence 136, App
42	454	9.7	760	15	US-10-177-253-136 Sequence 58, Appl
43	454	9.7	803	12	US-10-205-219-58 Sequence 43, Appl
44	431	9.2	803	12	US-10-205-219-58 Sequence 43, Appl
45	425.5	9.1	691	10	US-09-976-674-43 Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-09-976-674-1
Sequence 1, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akinsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junina, Jean-Louis
TITLE OR INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976, 674
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 882
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-674-1

Query Match	100.0%	Score 4700;	DB 10;	Length 882;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 882;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAAMMETEQGVETFTADG	ENIESQDRPKLEPFYVERYSMSQLKLLADTRKYGVNM	60
Db	1	MAAMMETEQGVETFTADG	ENIESQDRPKLEPFYVERYSMSQLKLLADTRKYGVNM	60
QY	61	AKRPHDPMTVKRNDPDPGHSRDRYYTAMSGENRENTLFTSEIPIKTIINRAAVMLSKPRL	120	
Db	61	AKRPHDPMTVKRNDPDPGHSRDRYYTAMSGENRENTLFTSEIPIKTIINRAAVMLSKPRL	120	
QY	121	DLFQATLDIGMSREBELLRERKRIGTVGIAVDYHQSGTFLFOAGSGIYHVMDGPGG	180	
Db	121	DLFQATLDIGMSREBELLRERKRIGTVGIAVDYHQSGTFLFOAGSGIYHVMDGPGG	180	

```

QY 161 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTREERRLTYVHNEL 240
Db 161 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTREERRLTYVHNEL 240
QY 241 ANMEEDASAGVATFVLOEFPDRYSGYWMCFAETTPSGKILRLIYENDESVEIITHV 300
Db 241 ANMEEDASAGVATFVLOEFPDRYSGYWMCFAETTPSGKILRLIYENDESVEIITHV 300
QY 301 TSPMLSTRADSFRRPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPELIFEGVE 360
Db 301 TSPMLSTRADSFRRPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPELIFEGVE 360
QY 361 YIARAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERORLIESVDSYPL 420
Db 361 YIARAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERORLIESVDSYPL 420
QY 421 ILYEETTDIMINIHDI FHVFPQSHHEBIEFI FASECKTGFRHLKYITSLKESKYKSSG 480
Db 421 ILYEETTDIMINIHDI FHVFPQSHHEBIEFI FASECKTGFRHLKYITSLKESKYKSSG 480
QY 481 GLPAPSDFKCPIKEEIALITSGEMVYLGHRGNSIQVDEVRLVYFEGTKDSPLEHLLYVS 540
Db 481 GLPAPSDFKCPIKEEIALITSGEMVYLGHRGNSIQVDEVRLVYFEGTKDSPLEHLLYVS 540
QY 541 YVNPGEVTRLTDRGYSHSCCISOHCDPFIISKYSNKNPCHVCVSLYKLSPEDDPTCKTKEF 600
Db 541 YVNPGEVTRLTDRGYSHSCCISOHCDPFIISKYSNKNPCHVCVSLYKLSPEDDPTCKTKEF 600
QY 601 WATILDSAGPLPDYTPPELIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFYGGPOVOL 660
Db 601 WATILDSAGPLPDYTPPELIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFYGGPOVOL 660
QY 661 VNNRFKGVKYPRLNTLASLGYVVVVYIDNRGSGCHRGKKEGAFKYMGOIEIDDOVEGLQY 720
Db 661 VNNRFKGVKYPRLNTLASLGYVVVVYIDNRGSGCHRGKKEGAFKYMGOIEIDDOVEGLQY 720
QY 721 LASRYDFIDLDVRVGHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYTERYM 780
Db 721 LASRYDFIDLDVRVGHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYTERYM 780
QY 781 GHPDNEQGYLYGSYVAMQAEKFPSEPNRLLLHGFIDENVHFAHTSILSLVRAKGYD 840
Db 781 GHPDNEQGYLYGSYVAMQAEKFPSEPNRLLLHGFIDENVHFAHTSILSLVRAKGYD 840
QY 841 LQIYPOERHSIRVPESGHEYLHLHYIQENIGSRIALKVI 882
Db 841 LQIYPOERHSIRVPESGHEYLHLHYIQENIGSRIALKVI 882

```

RESULT 2

```

US-10-054-776-2
; Sequence 2, Application US/10054776
; Publication No. US20030165818A1
; GENERAL INFORMATION:
; APPLICANT: Mark Robert Edbrooke
; APPLICANT: Alan Peter Lewis
; TITLE OF INVENTION: NOVEL PROTEIN
; FILE REFERENCE: QGI0420S
; CURRENT APPLICATION NUMBER: US/10/054.776
; CURRENT FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-776-2

```

```

Query Match 100.0%; Score 4700; DB 12; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MAAMTEBOLGYEIPETADCEENISODRPKLEPPYVERYSQJLKLADTRKHGYMM 60
Db 1 MAAMTEBOLGYEIPETADCEENISODRPKLEPPYVERYSQJLKLADTRKHGYMM 60
QY 61 AKAPDFMFKVKNDDPGPHSDRIYLLAMSGEENRENTLIFSEIPIKTIINBAALVLMISKPL 120
Db 61 AKAPDFMFKVKNDDPGPHSDRIYLLAMSGEENRENTLIFSEIPIKTIINBAALVLMISKPL 120
QY 121 DLFOATLDYGMYSREBELIREKRIIGTVGIASVDYHQSQTFLFOAGSGIYHVKGQFQG 180
Db 121 DLFOATLDYGMYSREBELIREKRIIGTVGIASVDYHQSQTFLFOAGSGIYHVKGQFQG 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTREERRLTYVHNEL 240
Db 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTREERRLTYVHNEL 240
QY 241 ANMEEDASAGVATFVLOEFPDRYSGYWMCFAETTPSGKILRLIYENDESVEIITHV 300
Db 241 ANMEEDASAGVATFVLOEFPDRYSGYWMCFAETTPSGKILRLIYENDESVEIITHV 300
QY 301 TSPMLSTRADSFRRPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPELIFEGVE 360
Db 301 TSPMLSTRADSFRRPKGTANPKYTFKMSIIMDAEGRIIDVIDKELIQPELIFEGVE 360
QY 361 YIARAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERORLIESVDSYPL 420
Db 361 YIARAGWTEGKYAMSILDRSQTRLQIVLISPELFIPEDDVMERORLIESVDSYPL 420
QY 421 ILYEETTDIMINIHDI FHVFPQSHHEBIEFI FASECKTGFRHLKYITSLKESKYKSSG 480
Db 421 ILYEETTDIMINIHDI FHVFPQSHHEBIEFI FASECKTGFRHLKYITSLKESKYKSSG 480
QY 481 GLPAPSDFKCPIKEEIALITSGEMVYLGHRGNSIQVDEVRLVYFEGTKDSPLEHLLYVS 540
Db 481 GLPAPSDFKCPIKEEIALITSGEMVYLGHRGNSIQVDEVRLVYFEGTKDSPLEHLLYVS 540
QY 541 YVNPGEVTRLTDRGYSHSCCISOHCDPFIISKYSNKNPCHVCVSLYKLSPEDDPTCKTKEF 600
Db 541 YVNPGEVTRLTDRGYSHSCCISOHCDPFIISKYSNKNPCHVCVSLYKLSPEDDPTCKTKEF 600
QY 601 WATILDSAGPLPDYTPPELIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFYGGPOVOL 660
Db 601 WATILDSAGPLPDYTPPELIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFYGGPOVOL 660
QY 661 VNNRFKGVKYPRLNTLASLGYVVVVYIDNRGSGCHRGKKEGAFKYMGOIEIDDOVEGLQY 720
Db 661 VNNRFKGVKYPRLNTLASLGYVVVVYIDNRGSGCHRGKKEGAFKYMGOIEIDDOVEGLQY 720
QY 721 LASRYDFIDLDVRVGHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYTERYM 780
Db 721 LASRYDFIDLDVRVGHGWSYGYLSLMAIMORSDFRVAIAGAPVTLMIFYDTGYTERYM 780
QY 781 GHPDNEQGYLYGSYVAMQAEKFPSEPNRLLLHGFIDENVHFAHTSILSLVRAKGYD 840
Db 781 GHPDNEQGYLYGSYVAMQAEKFPSEPNRLLLHGFIDENVHFAHTSILSLVRAKGYD 840
QY 841 LQIYPOERHSIRVPESGHEYLHLHYIQENIGSRIALKVI 882
Db 841 LQIYPOERHSIRVPESGHEYLHLHYIQENIGSRIALKVI 882

```

RESULT 3

```

US-10-170-789-38
; Sequence 38, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Welch, Nadine
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,

```



```

; SEQ ID NO 19
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-19

```

```

Query Match      74.6%; Score 3504; DB 10; Length 658;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MAAMETEQGLVEIETADCEENIESODRPKLEPPYVERYSQKLLADTRKHYGM 60
DB 1 MAAMETEQGLVEIETADCEENIESODRPKLEPPYVERYSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVKRNDPDGPHSDRIYYLAMS GENRENTLFYSELPKTIINAAVLMISWKPL 120
DB 61 AKAPHEMFVKRNDPDGPHSDRIYYLAMS GENRENTLFYSELPKTIINAAVLMISWKPL 120
QY 121 DLFOATLDYGMYSREBELRERKRI GTVGIASYDHQSGTFLFOAGSGIYHVKGDPG 180
DB 121 DLFOATLDYGMYSREBELRERKRI GTVGIASYDHQSGTFLFOAGSGIYHVKGDPG 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDIMISNIVTREERLTYVHNE 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDIMISNIVTREERLTYVHNE 240
QY 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEI 300
DB 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEI 300
QY 301 TSPMLERADSFYRPTGTANPKYTFKMSIIMIDAGR1IDVIDKELIQFEILLFEGVE 360
DB 301 TSPMLERADSFYRPTGTANPKYTFKMSIIMIDAGR1IDVIDKELIQFEILLFEGVE 360
QY 361 YIARAGWTPESKAYMSILLDRSQTRLOIVLISPELFIPEDDVNERQLISVDSVPL 420
DB 361 YIARAGWTPESKAYMSILLDRSQTRLOIVLISPELFIPEDDVNERQLISVDSVPL 420
QY 421 IIVETTDIMINHDIHFVPOSHHEIEFI FASECTGFPHLYKITSILKESYKSSG 480
DB 421 IIVETTDIMINHDIHFVPOSHHEIEFI FASECTGFPHLYKITSILKESYKSSG 480
QY 481 GLPAPSDPKPIKEIAITSGEWEVLGRHGSNIQVDEVRRLVYEGTDSPLEHLLYVS 540
DB 481 GLPAPSDPKPIKEIAITSGEWEVLGRHGSNIQVDEVRRLVYEGTDSPLEHLLYVS 540
QY 541 YVNPGEVTRLTDGRYSHSCCISOHCDPFISKYSNQKNPHCVSLYKLSPEDDPTCKTEF 600
DB 541 YVNPGEVTRLTDGRYSHSCCISOHCDPFISKYSNQKNPHCVSLYKLSPEDDPTCKTEF 600
QY 601 WATLLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGG 655
DB 601 WATLLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGG 655

```

RESULT 5

```

US-09-976-674-11
; Sequence 11, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976, 674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11

```

```

; LENGTH: 661
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-11

```

```

Query Match      74.6%; Score 3504; DB 10; Length 661;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MAAMETEQGLVEIETADCEENIESODRPKLEPPYVERYSQKLLADTRKHYGM 60
DB 1 MAAMETEQGLVEIETADCEENIESODRPKLEPPYVERYSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVKRNDPDGPHSDRIYYLAMS GENRENTLFYSELPKTIINAAVLMISWKPL 120
DB 61 AKAPHEMFVKRNDPDGPHSDRIYYLAMS GENRENTLFYSELPKTIINAAVLMISWKPL 120
QY 121 DLFOATLDYGMYSREBELRERKRI GTVGIASYDHQSGTFLFOAGSGIYHVKGDPG 180
DB 121 DLFOATLDYGMYSREBELRERKRI GTVGIASYDHQSGTFLFOAGSGIYHVKGDPG 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDIMISNIVTREERLTYVHNE 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDIMISNIVTREERLTYVHNE 240
QY 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEI 300
DB 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGKILRLIYENDESEVEI 300
QY 301 TSPMLERADSFYRPTGTANPKYTFKMSIIMIDAGR1IDVIDKELIQFEILLFEGVE 360
DB 301 TSPMLERADSFYRPTGTANPKYTFKMSIIMIDAGR1IDVIDKELIQFEILLFEGVE 360
QY 361 YIARAGWTPESKAYMSILLDRSQTRLOIVLISPELFIPEDDVNERQLISVDSVPL 420
DB 361 YIARAGWTPESKAYMSILLDRSQTRLOIVLISPELFIPEDDVNERQLISVDSVPL 420
QY 421 IIVETTDIMINHDIHFVPOSHHEIEFI FASECTGFPHLYKITSILKESYKSSG 480
DB 421 IIVETTDIMINHDIHFVPOSHHEIEFI FASECTGFPHLYKITSILKESYKSSG 480
QY 481 GLPAPSDPKPIKEIAITSGEWEVLGRHGSNIQVDEVRRLVYEGTDSPLEHLLYVS 540
DB 481 GLPAPSDPKPIKEIAITSGEWEVLGRHGSNIQVDEVRRLVYEGTDSPLEHLLYVS 540
QY 541 YVNPGEVTRLTDGRYSHSCCISOHCDPFISKYSNQKNPHCVSLYKLSPEDDPTCKTEF 600
DB 541 YVNPGEVTRLTDGRYSHSCCISOHCDPFISKYSNQKNPHCVSLYKLSPEDDPTCKTEF 600
QY 601 WATLLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGG 655
DB 601 WATLLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGG 655

```

RESULT 6

```

US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976, 674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 690

```

TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-674-7

Query Match 74.6%; Score 3504; DB 10; Length 690;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MAAMETEQLGVEIFETADCEENIESQDRPLKEFFYVERYSQKLLADTRKHYGM 60
DB 1 MAAMETEQLGVEIFETADCEENIESQDRPLKEFFYVERYSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVVRNDPDGPHSDRIYYLAMSNGENRENTLFYSEIPTINRAAVLMSWKPDL 120
DB 61 AKAPHEMFVVRNDPDGPHSDRIYYLAMSNGENRENTLFYSEIPTINRAAVLMSWKPDL 120
QY 121 DLFOATLDYGMYSSEBELRERKRGIVGASDYHQSSTFLQAGSGIHYVDGPGQ 180
DB 121 DLFOATLDYGMYSSEBELRERKRGIVGASDYHQSSTFLQAGSGIHYVDGPGQ 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPWIATIHNSNDIWSNIVTREERLTYVHNE 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPWIATIHNSNDIWSNIVTREERLTYVHNE 240
QY 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTSQKILRLIYENDESEVEIHHV 300
DB 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTSQKILRLIYENDESEVEIHHV 300
QY 301 TSPMLERRADSRFPYPTGTANPKVTKMSEIMDAGR11DVIDKELIQPELLEFEGVE 360
DB 301 TSPMLERRADSRFPYPTGTANPKVTKMSEIMDAGR11DVIDKELIQPELLEFEGVE 360
QY 361 YIARAGTPEGKYAMSLILDRSQTRQIVLISELFIPEVDDVWERORLIESVDSVTP 420
DB 361 YIARAGTPEGKYAMSLILDRSQTRQIVLISELFIPEVDDVWERORLIESVDSVTP 420
QY 421 IYEEETDWINIHDFHVPQSHHEEIEFIFASECKTGRRLYKITSILKESYKSSG 480
DB 421 IYEEETDWINIHDFHVPQSHHEEIEFIFASECKTGRRLYKITSILKESYKSSG 480
QY 481 GLPAPSDFKPIKEEIAITSGEWEVIGRHGSIQVDEVRLLVFEETKDSPLEHHLVVS 540
DB 481 GLPAPSDFKPIKEEIAITSGEWEVIGRHGSIQVDEVRLLVFEETKDSPLEHHLVVS 540
QY 541 YVNPGEVTRLTDGYSHSCISQCHDFISKYSNQNPKHCVSLYKSSPEDDPCTKKEF 600
DB 541 YVNPGEVTRLTDGYSHSCISQCHDFISKYSNQNPKHCVSLYKSSPEDDPCTKKEF 600
QY 601 MATILDSAGLPDPTPELIFSFESTTGFTLYGMLYKPHDLPQKXPTVLFYVG 655
DB 601 MATILDSAGLPDPTPELIFSFESTTGFTLYGMLYKPHDLPQKXPTVLFYVG 655
```

RESULT 7
US-09-976-674-21
Sequence 21, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akimsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 21
LENGTH: 613
TYPE: PRT
ORGANISM: Homo sapiens

ORGANISM: Homo sapiens
US-09-976-674-21

Query Match 68.9%; Score 3236; DB 10; Length 613;
Best Local Similarity 100.0%; Pred. No. 1,4e-301;
Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MAAMETEQLGVEIFETADCEENIESQDRPLKEFFYVERYSQKLLADTRKHYGM 60
DB 1 MAAMETEQLGVEIFETADCEENIESQDRPLKEFFYVERYSQKLLADTRKHYGM 60
QY 61 AKAPHEMFVVRNDPDGPHSDRIYYLAMSNGENRENTLFYSEIPTINRAAVLMSWKPDL 120
DB 61 AKAPHEMFVVRNDPDGPHSDRIYYLAMSNGENRENTLFYSEIPTINRAAVLMSWKPDL 120
QY 121 DLFOATLDYGMYSSEBELRERKRGIVGASDYHQSSTFLQAGSGIHYVDGPGQ 180
DB 121 DLFOATLDYGMYSSEBELRERKRGIVGASDYHQSSTFLQAGSGIHYVDGPGQ 180
QY 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPWIATIHNSNDIWSNIVTREERLTYVHNE 240
DB 181 FTQOPLRPNLVETSCPNIRMDPKLCPADPWIATIHNSNDIWSNIVTREERLTYVHNE 240
QY 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTSQKILRLIYENDESEVEIHHV 300
DB 241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTSQKILRLIYENDESEVEIHHV 300
QY 301 TSPMLERRADSRFPYPTGTANPKVTKMSEIMDAGR11DVIDKELIQPELLEFEGVE 360
DB 301 TSPMLERRADSRFPYPTGTANPKVTKMSEIMDAGR11DVIDKELIQPELLEFEGVE 360
QY 361 YIARAGTPEGKYAMSLILDRSQTRQIVLISELFIPEVDDVWERORLIESVDSVTP 420
DB 361 YIARAGTPEGKYAMSLILDRSQTRQIVLISELFIPEVDDVWERORLIESVDSVTP 420
QY 421 IYEEETDWINIHDFHVPQSHHEEIEFIFASECKTGRRLYKITSILKESYKSSG 480
DB 421 IYEEETDWINIHDFHVPQSHHEEIEFIFASECKTGRRLYKITSILKESYKSSG 480
QY 481 GLPAPSDFKPIKEEIAITSGEWEVIGRHGSIQVDEVRLLVFEETKDSPLEHHLVVS 540
DB 481 GLPAPSDFKPIKEEIAITSGEWEVIGRHGSIQVDEVRLLVFEETKDSPLEHHLVVS 540
QY 541 YVNPGEVTRLTDGYSHSCISQCHDFISKYSNQNPKHCVSLYKSSPEDDPCTKKEF 600
DB 541 YVNPGEVTRLTDGYSHSCISQCHDFISKYSNQNPKHCVSLYKSSPEDDPCTKKEF 600
QY 601 MATILDS 607
DB 601 MATILDS 607
```

RESULT 8
US-09-976-674-3
Sequence 3, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akimsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 863
TYPE: PRT
ORGANISM: Homo sapiens

US-09-976-674-3

Query Match	61.1%;	Score 2870;	DB 10;	Length 863;
Best Local Similarity	61.5%;	Pred. No. 3.7e-266;		
Matches 517;	Conservative 134;	Mismatches 187;	Indels 2;	Gaps 2

QY	35	YVVERYSOLKLLADTRKXHYMMAAAPHMFVXKNDPDGHSDBIYYIAMSGENRE	94
Db	24	FQVQKHSIDGKRSITIHGSRKTSGLIWNKAPHPQFQVQTDGSGPSHSLYYIAGMYGPRE	83
QY	95	NLFYSEIPIKTNRAAVLMSKPKLLDFOATLDGMYSRBEELLREKRKIGTVIASYD	154
Db	84	NSLIYSEIPKVKREALLLSWKQMLDHFQATPHHGVYSREBEILLREKKLGVEBITSYD	143
QY	155	YHOSSGTFLLPAGSGIYHVXKGGFOGFLQOPLRPNLVTSQCNIRMDPKLCPADPDWIAF	214
Db	144	FHSBSGFLFQASMSLFHCRDGGKXGFVWSPMKPLEITQCSGPMDEKICPADAPFSEF	203
QY	215	IHSNDIWMISNIVTREERLTYVHNHNAMEEDASAGAVATVLOEPPRYSYMMWCPAE	274
Db	204	INNSDLWVANIETGERRLTCHQGLSNVLDPPSAGAVATVLOEPPRFGYWWCPAAS	263
QY	275	TTPSGG-KILNLIYEENDESEVEIITHVSPMLETRRADSFRRPKGTANPYVTFMSSEIM	333
Db	264	WESSEGGKTLIRLYEEVDESEVEYIHWSPSPALEEKTDYSYRPRGSKNPKIALKLAHQ	323
QY	334	IDAGSRIIDVTDKLILOPELLFSGVEYIAAGMTPECKYMSILDBSSQRLQIVLSP	393
Db	324	TDGSGKIVSTOEKLVOPFSSLFPRVEYIAAGMTROCKYAMAFILDPQOMLOLVLPP	383
QY	394	ELFIFVEDVWERQRLIESVPDSTPLIYEETDIWMIHIDIFHFVPOSH-KEIEIFIF	452
Db	384	ALFIPSTENEBQRLASARAVPRNVQPYVVEYEVNWMINVAHDYPPFQSGDEBELCLR	443
QY	453	ASEKTEGFRHLKYKTLILKESKYYRSSGCLPABDFKPCIKYEBIATTSGENEVLGRHGSN	512
Db	444	ANECKTGFCHLYKTAVALKSGQYMSSEPFSGDEBFKPCIKYEBIATTSGENEVLARHGSK	503
QY	513	IQVDEVBRLYVEEETKDSPLEHNHLYVSVYAPGVTBLTDGVSHCISIOHCPFIISKY	572
Db	504	IWNNEFKLYIFQGTRODPLEHNHLYVSYAAGIVRLITTPGFHSSCSMSQNFPMFVSHY	563
QY	573	SNQKNPCVSLYKLSPEDEPTCKTEKFMAITLDSAGPLDPYTPPELIFSFESTGFTLYG	632
Db	564	SSVSTPECVHYKLSGDDDBLHKOPRFWMSMMAASCPDYPVPELPHHTRSDVXLHG	623
QY	633	MLVKPHDLOGKKYPTLFTYGGQOULVNNRPFQVYKPYFRINTYASIGYVVVVIDNRGSC	692
Db	624	MYKPFALOGKKHPVLFWYGGQOVLVNNSFQIGITYLRINTLASIGYAVVVIDGSGSC	683
QY	693	HRLGKFGAERYXMGQIEIDDOVEGLOYLASRYOFLIDRGIHGSYGGGLSIALMAQR	752
Db	684	QRGRFEGALKMGQGYEIEDQVEGLOFVAEKYCFIDLSKVALHGSYGGGLSIMGLIHK	743
QY	753	SDIFRVALAGPYVTLMYDTGYTERYMGHPDQNEOGYILGSYVMAQAEKPSSENNRLLL	812
Db	744	POVFKAVALAGPYVWMAVDTGYTERYMDVPENNQOHEYAGSVVALHVEKLPNEENRLLIL	803
QY	813	HGGLDEVENFAHYSILSFLVRAKGPVDLOIYQOERHSIVPREGERYELHLHYLOENT	872
Db	804	HGGLDEVENFAHYNFLVSLIRKGPYQLOIYPERKHSICPSGGEHYEVTLLHFOEVL	863

RESULT 9
US-09-976-674-23

```

; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/376,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 892
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-376-674-23

```

Query Match	61.1%;	Score 2870;	DB 10;	Length 892;
Best Local Similarity	61.5%;	Pred. No. 4e-266;		
Matches 517;	Conservative 134;	Mismatches 187;	Indels 2;	Gaps 2

QY	35	YVYRYSWSQJUKLLADTRKXHGIMMAKAPHDPMFYKRNDDGDGSHDPIYTLANSGENRE	94
Db	53	FOVKSHSMDGJLRSTIHSSRKYSGLIYNKAPHPDFOVYKTDBSGSHSLYLYIMPGYGRE	112
QY	95	NLFYSBIPKTIKINAAVLM,SWKPLDL,FOALTLOYGMSREELLREKRIGTGVASIDY	154
Db	113	NSLILYSBIPKVRKEALLLLSWKQMDHFOQTPHHGYVSRBEELLREKRGVRSITSYD	172
QY	155	YHOSSGTFLFOAGSGIYHVKDGPQGTQOPLRPNLVETS,CPNIRMDKLCPADPDWIAF	214
Db	173	FHSBSGFELFOASNSLPHCRDGGKGNQGMVSMFKLEIKTQCSGCPMRPKICPADPAFRSF	232
QY	215	IHSNDIMSNVTVEERRLTYVNHLEINMEDASAGVATVLOEPPFRYGGYWKCPRAE	274
Db	233	INNDDLWAVANIETGEERRLTFCHOGLSNVLDDPSAGAVATVIOEBPRFLGYWKCPRAE	292
QY	275	TTBSGG-KILRIIYEENDESEVELIHVTSPL,ETRRADSFRRPYKGTANPYVTEKMEIM	333
Db	293	WEGSEGGKTLRIIYEVEDESEVEYIHPSPALBEKTKDSYRPTGSKNPKIALKLMFQ	352
QY	334	IDAEGRIIDVIDELLTOPPELLPBGVYIARAGCTPBGKTA,MSILLDSQTRLOIVLISP	393
Db	353	TDGSGKIVTSQOEKELVQPSLSLPPKVEYIARAGCTBKGKAMAMFELDRPOOMLOTLVLBP	412
QY	394	ELFIPVDDMYERORLLIESVPDSTPLIIYEETDIIINIHDPHYPOSH-BEIEBIF	452
Db	413	ALFIPSTENEGOLHASARAVPRNVQPVVYEEVTNWINVHDLFYPPPOSEGEDBELCFLR	472
QY	453	ASBCKTGFRLUYITSLIKESKYYKSSGGLPABSDFKCPIKEBIA,ITSGEMEVLGRHGSN	512
Db	473	ANBCKTGFCMLYKTVATVLKSGQYDWSEPFSGDEDFCPIKEBIALITSGEMEVLARHGSK	532
QY	513	IQVDEYBRILVFPBGTDSPLEHHLVYVSYNRPBVTRLTRGRVSHSCCIIQHCDFPISKY	572
Db	533	IMVNEEKILVYFQGTQDTPLEHHLVYVSYAAGEIYKLTTPGSHSOSMSQNFPMFVSHY	592
QY	573	SNQKNPHCVSLYKLSPEBDDPTCKTKEFWATILLDSAGPLDCDYPPPELTFSESTGFLYLG	632
Db	593	SSVSTPCVHYVLTGSDDDPLHKQRPFRMASMMEMASCPDYVPELPHHHTHSDVRLYG	652
QY	633	MLYKPHDLOPGKCYPTVLYFYGGPQVQLVNNRFKGYKRYFLNLTASTIGYVVVVIDNRGSC	692
Db	653	MYKPHALDPOGKGPVLYFYGGPQVQLVNNRSFKGIKYLRLNLTASTIGYVVVVIDNRGSC	712
QY	693	HRLGKEGAPKYMGGOIEIDDOVEG,GOYLASRYDFILDRVGHGMSYGGYSLIMLMOR	752
Db	713	QRLRAREGALKRQMGQVEIEDOVBG,GOFAEAKGCFIDLSRVALIHGMSYGGFSLMGLIHK	772
QY	753	SDIFRVALIAGAPVTLMIFYDTGYTEREYMGHPDQNEGOYLYGSVAMOAEEKFPSEPNLLIL	812
Db	773	PQVFKVALIAGAPVTLVMMAYDGTGYTERYMDVPERNNQGYEAGSVALAHVEKLPNEPNLLIL	832
QY	813	HGGLDENHFAHRSIILSLFVAGKRYD,CIYQOEHSIRVPSGSHYELHLIHYIOENL	872
Db	833	HGFLDENHFAHRSIILSLFVAGKRYD,CIYQOEHSIRVPSGSHYELHLIHYIOENL	892

RESULT 10
US-09-976-674-27
; Sequence 27, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 892
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-976-674-27

Query Match 61.1%; Score 2870; DB 10; Length 892;
Best Local Similarity 61.5%; Pred. No. 4e-266;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;
QY 35 FVYERYSMQKLKLLADTRKHYGMMAKAPHDPMFYKRNDDPGPHSDRIYIYLAAMSGENE 94
DB 53 FOVQKSHMDGLRSIIHSGSRKYSGLIVNKAHPDFOVQKDESGPHSHRIYIYLAAMPYGSRE 112
QY 95 NTLFYSEIPKTIINRAVAVLMSKPLLDLFOATLDYGMYSREBELLRERKRIGVGTIASYD 154
DB 113 NSLTYSEIPKTKRKALLLSWKQMLDHPQATPHHGVSREBELLRERKRIGVGTITSYD 172
QY 155 YHOGSGTFLFOAGSGIYHVKGPGQFTQOPLRPNLVETSCPNIRMDPKLCPADPDMIAF 214
DB 173 FHSSEGLFLFOASNSLIFHCRODGKNGFMVSPKPLEIKITQCGSPRMDPKICPADPAFSPF 232
QY 215 IHSNDIWSNIVTREERRLTYVHNELANNEEDARASAGVATFVLOEFDRYSGYMWCPKAE 274
DB 233 INNSDLWVANILETGERRLTFCHQGLSNVLDPKSAGVATFVIOEFDRFTGYMWCPKAS 292
QY 275 TTPSGG-KILRLIYENDESEVEIIVHTSPMLSTRADFRPKGTANPKVTFPKMSEIM 333
DB 293 WEGSGGLKTLRLIYEVDESEVEIVHVPSPALEERKTSYRPRKGSKNPKITALKLAEPQ 352
QY 334 IDAEGRIIDVIDKELIOPELIFEGVEYIARAGMTPEKGYAMSIILDSQTRLQIVLISP 393
DB 353 TDSQGIKIVSTQEKELVQPFSSLPFVVEYIARAGMTRODKIYAMWFLDRPQWLQIVLISP 412
QY 394 ELFIPEDDVNERQRLIESVPDSVTPPLIYEETDIDIMINIDIFHVFQSH-EEIEIFTF 452
DB 413 ALFIPTSTNEBQRLASARAVERNQPVYVEVTVWVWVNHVIDIFPPQSGEEDLCLFLR 472
QY 453 ASECKTGRRLIYKTSILKESKYRSSGGLPAPSDFKPICKEIALITGEMEVLCGRHNSN 512
DB 473 ANECKTGFCHLYKTAVLKSQGYDMSBPSPREDEFFKCIKEIALITGEMEVLCARHNSK 532
QY 513 IQVDEVRRLVYFEGTKDSPLEHHLIYVSYVNGEVATRLTDRGYSHSCCISQCHDFFIISKY 572
DB 533 IWNVEETKLYVFQGTOKDPLRLEHLLIYVSYEAAGBIIVRLITTPGFSSCSMSQNFDMFVSHY 592
QY 573 SNQKNPHVSLIYKLSPPDDPCTCKKEWATITLDSAGLPRYTPPEIFSFSTGTFTLYG 632
DB 593 SSVSTPVCVHYKLSGPDDELPHKQPRFWMASMEALASCPDYVPEIIFHFTRSQVRLYG 652
QY 633 MLYEPHDQPGKKTPTVLFYIGPQVOLLWNRFKGVKFRINTLASLGYVVVVVDNRGSC 692
DB 653 MIYEPHALQPGKKTPTVLFYIGPQVOLLWNSFKGIKTLRINTLASLGYAVVVDNRGSC 712
QY 693 HRLKFGCAFKYKMGQIEIDDQVEGLQYLAIRYDFIDLRVGIHGWSYGYLSIMALMQR 752

DB 713 QRGIRREGALKNMQGVQIEIDQVEGLQFVAEKYGFIDLSRAVJHGWSYGYLSIMGLIHK 772
QY 753 SDIFRYALAGAPVIMTFVDTGTERYMGHPDQEGCYIYGSVAMAKESPSENNLL 812
DB 773 POFKVALIAGAPVIMWADYGTERTMDVBNQHGVEAGSVALHEKLPNEBNRLIL 832
QY 813 HGFLENVHFAHTSILSFLVRAGKPYDLOIYPOERSIRIVSPSGHEYLHLHYLOENL 872
DB 833 HGFLENVHFAHTSILSFLVRAGKPYDLOIYPOERSIRIVSPSGHEYLHLHYLOENL 892

RESULT 11
US-09-976-674-33
; Sequence 33, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 879
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-976-674-33

Query Match 60.0%; Score 2820.5; DB 10; Length 879;
Best Local Similarity 60.7%; Pred. No. 2.2e-261;
Matches 510; Conservative 132; Mismatches 183; Indels 15; Gaps 3;
QY 35 FVYERYSMQKLKLLADTRKHYGMMAKAPHDPMFYKRNDDPGPHSDRIYIYLAAMSGENE 94
DB 53 FOVQKSHMDGLRSIIHSGSRKYSGLIVNKAHPDFOVQKDESGPHSHRIYIYLAAMPYGSRE 112
QY 95 NTLFYSEIPKTIINRAVAVLMSKPLLDLFOATLDYGMYSREBELLRERKRIGVGTIASYD 154
DB 113 NSLTYSEIPKTKRKALLLSWKQMLDHPQATPHHGVSREBELLRERKRIGVGTITSYD 172
QY 155 YHOGSGTFLFOAGSGIYHVKGPGQFTQOPLRPNLVETSCPNIRMDPKLCPADPDMIAF 214
DB 173 FHSSEGLFLFOASNSLIFHCRODGKNGFMVSPKPLEIKITQCGSPRMDPKICPADPAFSPF 232
QY 215 IHSNDIWSNIVTREERRLTYVHNELANNEEDARASAGVATFVLOEFDRYSGYMWCPKAE 274
DB 233 INNSDLWVANILETGERRLTFCHQGLSNVLDPKSAGVATFVIOEFDRFTGYMWCPKAS 292
QY 275 TTPSGG-KILRLIYENDESEVEIIVHTSPMLSTRADFRPKGTANPKVTFPKMSEIM 333
DB 293 WEGSGGLKTLRLIYEVDESEVEIVHVPSPALEERKTSYRPRKGSKNPKITALKLAEPQ 352
QY 334 IDAEGRIIDVIDKELIOPELIFEGVEYIARAGMTPEKGYAMSIILDSQTRLQIVLISP 393
DB 353 TDSQGIKIVSTQEKELVQPFSSLPFVVEYIARAGMTRODKIYAMWFLDRPQWLQIVLISP 412
QY 394 ELFIPEDDVNERQRLIESVPDSVTPPLIYEETDIDIMINIDIFHVFQSH-EEIEIFTF 452
DB 413 ALFIPTSTNEBQRLASARAVERNQPVYVEVTVWVWVNHVIDIFPPQSGEEDLCLFLR 472
QY 453 ASECKTGRRLIYKTSILKESKYRSSGGLPAPSDFKPICKEIALITGEMEVLCGRHNSN 512
DB 473 ANECKTGFCHLYKTAVLKSQGYDMSBPSPREDEFFKCIKEIALITGEMEVLCARHNS- 531
QY 513 IQVDEVRRLVYFEGTKDSPLEHHLIYVSYVNGEVATRLTDRGYSHSCCISQCHDFFIISKY 572

Db 532 -----KGTQDTPLEHNLVYVSYEAGEIVRLTTPGFSHSCSMQNFDMFVSHY 579

QY 573 SNQKPHCVSLYKLSSEDPDPTCKTEKFWATILDSAGLPDYTPPEIFSESTGFTLYG 632

Db 580 SSVSTPCVHYKLSGDDPLHKQPRFMAEMEAASCPDYVPELFFHFRTRDVLHG 639

QY 633 MLKYPHDLQPKKKPTVLYFGGPOVQVYNNRFGVKYFRPLNTLASIGYVVVVDNRGSC 692

Db 640 MYKPHALQPKKHPVLYFGGPOVQVYNNRFGVKYFRPLNTLASIGYVVVVDNRGSC 699

QY 693 HRGLKFEAGAFKYMKGQLEIDDVEGLQYASRYDFILDRYGHGWSYGGYLSMALMQR 752

Db 700 QRGIRFEGALKMKGQVEIEDVGLQVFAEKYFIDLSRYALHGWSYGGYLSMALMQR 759

QY 753 SDIFRYAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVYAMOAEKFPSEPNRLILL 812

Db 760 POFKVAIAAGAPVTLVMAVDTGYTERYMDVPENNQHGEGVYALHVEKLPNEPNRLILL 819

QY 813 HGFLDENVHFAHNSILSFLVRAGKPYDLOIYQERHSIRVPSGHEHYELHLHYOLENT 872

Db 820 HGFLDENVHFAHNSILSFLVRAGKPYDLOIYQERHSIRVPSGHEHYELHLHYOLENT 879

RESULT 12

US-09-976-674-35

Sequence 35, Application US/09976674

Patent No. US20020115843A1

GENERAL INFORMATION:

APPLICANT: Qi, Steve

APPLICANT: Akinsanya, Karen

APPLICANT: Riviere, Pierre

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DBPIV

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976,674

CURRENT FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,117

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 61

SOFTWARE: Patentin version 3.1

SEQ ID NO 35

LENGTH: 879

TYPE: PRT

ORGANISM: Homo sapiens

US-09-976-674-35

Query Match 60.0%; Score 2820.5; DB 10; Length 879;

Best Local Similarity 60.7%; Pred. No. 2.2e-261;

Matches 510; Conservative 132; Mismatches 183; Indels 15; Gaps 3;

QY 35 FVYERYWSQKLKLLADTRKYHGYMAKAPHDMFVKRNDPDGPHSDRYIYLAWSGENRE 94

Db 53 FOYQKSMWGLKRSITIHGSRKYSGLIVNKAAPHDFQVQKTDSESPHSHRLYYLGMPPGSR 112

QY 95 NTLFYSEIPKTIINRAAVLMLSWKPLLDLFOATLDYGYMSREBELLRERRRIGTVGIASYD 154

Db 113 NSLIVSEIPKTKYRKALLLSWKQMLDHFQATPHHGYVSRBELLRERRRIGTVGIASYD 172

QY 155 YHOGSGTFLFOAGSGIYHVKGPGFTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214

Db 173 FHSESGTLFLFOASNSLFFHCRDGGKNGFVMSPMKPLEIKTQCSGPRMDPKICPADPAFSPF 232

QY 215 HSNIDWISNITVTRERRRLTYVHNELANNEEDASGVATFVLOEFDFISYGVWCPKAE 274

Db 233 INNSDLWVANITGERRRLTFHOGISNVLDDPKSGAVATFVLOEFDFISYGVWCPKAE 292

QY 275 TTPSGG-KILRLIYEENDESEVEIHWTSPLMLETREDSFRYPKGTGTANPKYTFPMSEIM 333

Db 293 WEGSGELKTLRIIYEVDESEVEIHWTSPLMLETREDSFRYPKGTGTANPKYTFPMSEIM 352

QY 334 IDABRIIDVIDKELIOPPELIFEGYVYIARAGWPEGKYANSILDSQTLQIVLISP 393

Db 353 TDSQGIIVSTQEKELVQPPSSLFPKVEYIARAGWTRDGYAWAMFLDRQOMLQIVLISP 412

QY 394 ELFIYEDDVMEQORQLIESYVDSTPLIYEETDLMINIHDI PHVEPOSH -EEIEIFIF 452

Db 413 ALPISTENEEQRLASRAVARNQYVYVEETVWINVHDIYFPPOSEGEDELQFLR 472

QY 453 ASECKTGFRHLKYITSILKESKYRSSGGAPASDFKCPYKEBIAITSGMEVYGRGSGN 512

Db 473 ANECKTGFCALYKVTAVLKSQGYDMSPEFSGDEBFKCPYKEBIAITSGMEVYGRGSGN 531

QY 513 IQDVEARLVYFEGTDSPLLEHNLVYVSYNPGEVTRLTDRGYSHSCCISQHCDFPISKY 572

Db 532 -----KGTQDTPLEHNLVYVSYEAGEIVRLTTPGFSHSCSMQNFDMFVSHY 579

QY 573 SNQKPHCVSLYKLSSEDPDPTCKTEKFWATILDSAGLPDYTPPEIFSESTGFTLYG 632

Db 580 SSVSTPCVHYKLSGDDPLHKQPRFMAEMEAASCPDYVPELFFHFRTRDVLHG 639

QY 633 MLKYPHDLQPKKKPTVLYFGGPOVQVYNNRFGVKYFRPLNTLASIGYVVVVDNRGSC 692

Db 640 MYKPHALQPKKHPVLYFGGPOVQVYNNRFGVKYFRPLNTLASIGYVVVVDNRGSC 699

QY 693 HRGLKFEAGAFKYMKGQLEIDDVEGLQYASRYDFILDRYGHGWSYGGYLSMALMQR 752

Db 700 QRGIRFEGALKMKGQVEIEDVGLQVFAEKYFIDLSRYALHGWSYGGYLSMALMQR 759

QY 753 SDIFRYAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVYAMOAEKFPSEPNRLILL 812

Db 760 POFKVAIAAGAPVTLVMAVDTGYTERYMDVPENNQHGEGVYALHVEKLPNEPNRLILL 819

QY 813 HGFLDENVHFAHNSILSFLVRAGKPYDLOIYQERHSIRVPSGHEHYELHLHYOLENT 872

Db 820 HGFLDENVHFAHNSILSFLVRAGKPYDLOIYQERHSIRVPSGHEHYELHLHYOLENT 879

RESULT 13

US-09-976-674-29

Sequence 29, Application US/09976674

Patent No. US20020115843A1

GENERAL INFORMATION:

APPLICANT: Qi, Steve

APPLICANT: Akinsanya, Karen

APPLICANT: Riviere, Pierre

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DBPIV

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976,674

CURRENT FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,117

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 61

SOFTWARE: Patentin version 3.1

SEQ ID NO 29

LENGTH: 832

TYPE: PRT

ORGANISM: Homo sapiens

US-09-976-674-29

Query Match 51.2%; Score 2406; DB 10; Length 832;

Best Local Similarity 59.0%; Pred. No. 1.4e-221;

Matches 434; Conservative 122; Mismatches 177; Indels 2; Gaps 2;

QY 35 FVYERYWSQKLKLLADTRKYHGYMAKAPHDMFVKRNDPDGPHSDRYIYLAWSGENRE 94

Db 53 FOYQKSMWGLKRSITIHGSRKYSGLIVNKAAPHDFQVQKTDSESPHSHRLYYLGMPPGSR 112

QY 95 NTLFYSEIPKTIINRAAVLMLSWKPLLDLFOATLDYGYMSREBELLRERRRIGTVGIASYD 154

Db 113 NSLIVSEIPKTKYRKALLLSWKQMLDHFQATPHHGYVSRBELLRERRRIGTVGIASYD 172

QY 155 YHOGSGTFLFOAGSGIYHVKGPGFTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214

Db 173 FHSESGTLFLFOASNSLFFHCRDGGKNGFVMSPMKPLEIKTQCSGPRMDPKICPADPAFSPF 232

QY 215 IHSNDIWSNVTREBRRLTYVHNELAMMEDASAGVATFVLQESBDRYSGYWCPCRAE 274
 Db 233 INNSDLWNAITGEERLITFCOGLSNVLDPKSGAVATFVQESBDRFTGVMWCPITAS 292
 QY 275 TTPSGG-KILRIIYENDESEVEIIHVSFMLETRRADSPRYPKGTANPKVFKMSEIM 333
 Db 293 WEGSEGLKTLRIIYEVEDESEVEIIVHPSPALBEKRTDSYRPRTSKPKIKLXLAERQ 352
 QY 334 IDAEGRIIDVIDKEILOPPELLFEGVEYIARAGMTBEGKAWNSILLDRSOTRLQIVLISP 393
 Db 353 TDSOGKIVSTQEKELVQFSSLPFKVEYIARAGMTBEGKAWNSILLDRSOTRLQIVLISP 412
 QY 394 ELFIPEVDVMEORQLIESVPSVTPPLIYEETDWINIHDIHFHVPQSH-EEIEFTF 452
 Db 413 ALFIPTSTENEOQLASARAVPRNQPYVVEVTNWVINDIYFPFQSGEDELCTLR 472
 QY 453 ASECKTGFRLHYKITSILKESKYYRSGGLPAPSDFKCPIKEBIAITSGMEVLAGHGSN 512
 Db 473 ANECKTGFCMLYKVTAVLKSQGYDWSPEFSGDEBFKCPIKEBIAITSGMEVLAGHGSN 532
 QY 513 IOVDEVRLVYFEGTKDPLSHHLVYVSVNPGEVTRLTDGYSHSCCISQHCDFISKY 572
 Db 533 IWNVEETLVYFQGTDPLEHHLVYVSEAGIYVRLITPGFSHSCMSQNFDFVSHY 592
 QY 573 SNOKNPHCVSLYKLSPEEDPTCKTKEFWATILDSAGPLDYPPEIFSFESTTGFTLYG 632
 Db 593 SSVSTPCVHYKLSGPDDDLHMQPRFMAEMEAASCPDYVPEIFHFTRSQVRLYG 652
 QY 633 MLKPHDLOPGKRYPTVLYFGGPOVULVNNRFGKVFYFNTLASLGYVAVVINDRSGC 692
 Db 653 MLYKPHALOFGKHPVLYFGGPOVULVNNRFGKIKYLRNTLASLGYAVVINDRSGC 712
 QY 693 HRGLKEGAFKRYKMGQIIEIDQVEGLQYLAARYDFIDLRVGIHWSYGYLSLMAIMOR 752
 Db 713 QRGIRFEGALKNMQGVIEIDQVEGLQYLAARYDFIDLRVGIHWSYGYLSLMAIMOR 772
 QY 753 SDIFRVAIAGAPVTL 767
 Db 773 PQVFKAOPLAAPPRL 787
 RESULT 14
 US-09-976-674-31
 ; Sequence 31, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akusanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Junien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: US 60/240,117
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 31
 ; LENGTH: 832
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-976-674-31

Query Match 51.2%; Score 2406; DB 10; Length 832;
 Best Local Similarity 59.0%; Pred. No. 1,4e-221;
 Matches 434; Conservative 122; Mismatches 177; Indels 2; Gaps 2;
 QY 35 FYVERYSQULKLLADTRKYHGVMMAKAPDPMFVGRNDPDPGSDRIYIYLMGSENE 94
 Db 53 FOYQKSHMDGRLSRIHSGRKYSGIIVAKAPDFQVQKDBSGPHRLIYILMGPSGRE 112
 QY 95 NTLFYSEIPTINPAVALMSKFLDLDFOATIDYGMYSREBELRERKRIGTVGIASVD 154

Db 113 NSLIYSEIPKVRKREALLLSMKOMLDFQATLPHHGYSRREELRERKRIGTVGITSYD 172
 QY 155 YHQSSTFLFOAGSGIYHVKDQPOGFTQDPLRPNLVETSCPNIRMDPKLCPADPMIAF 214
 Db 173 FHSSEGLFLOASNLIFCRDQKNGFVNSPMKLEIKTQCSGRMPKICPADPAFFSF 232
 QY 215 IHSNDIWSNVTREBRRLTYVHNELAMMEDASAGVATFVLQESBDRYSGYWCPCRAE 274
 Db 233 INNSDLWNAITGEERLITFCOGLSNVLDPKSGAVATFVQESBDRFTGVMWCPITAS 292
 QY 275 TTPSGG-KILRIIYENDESEVEIIHVSFMLETRRADSPRYPKGTANPKVFKMSEIM 333
 Db 293 WEGSEGLKTLRIIYEVEDESEVEIIVHPSPALBEKRTDSYRPRTSKPKIKLXLAERQ 352
 QY 334 IDAEGRIIDVIDKEILOPPELLFEGVEYIARAGMTBEGKAWNSILLDRSOTRLQIVLISP 393
 Db 353 TDSOGKIVSTQEKELVQFSSLPFKVEYIARAGMTBEGKAWNSILLDRSOTRLQIVLISP 412
 QY 394 ELFIPEVDVMEORQLIESVPSVTPPLIYEETDWINIHDIHFHVPQSH-EEIEFTF 452
 Db 413 ALFIPTSTENEOQLASARAVPRNQPYVVEVTNWVINDIYFPFQSGEDELCTLR 472
 QY 453 ASECKTGFRLHYKITSILKESKYYRSGGLPAPSDFKCPIKEBIAITSGMEVLAGHGSN 512
 Db 473 ANECKTGFCMLYKVTAVLKSQGYDWSPEFSGDEBFKCPIKEBIAITSGMEVLAGHGSN 532
 QY 513 IOVDEVRLVYFEGTKDPLSHHLVYVSVNPGEVTRLTDGYSHSCCISQHCDFISKY 572
 Db 533 IWNVEETLVYFQGTDPLEHHLVYVSEAGIYVRLITPGFSHSCMSQNFDFVSHY 592
 QY 573 SNOKNPHCVSLYKLSPEEDPTCKTKEFWATILDSAGPLDYPPEIFSFESTTGFTLYG 632
 Db 593 SSVSTPCVHYKLSGPDDDLHMQPRFMAEMEAASCPDYVPEIFHFTRSQVRLYG 652
 QY 633 MLKPHDLOPGKRYPTVLYFGGPOVULVNNRFGKVFYFNTLASLGYVAVVINDRSGC 692
 Db 653 MLYKPHALOFGKHPVLYFGGPOVULVNNRFGKIKYLRNTLASLGYAVVINDRSGC 712
 QY 693 HRGLKEGAFKRYKMGQIIEIDQVEGLQYLAARYDFIDLRVGIHWSYGYLSLMAIMOR 752
 Db 713 QRGIRFEGALKNMQGVIEIDQVEGLQYLAARYDFIDLRVGIHWSYGYLSLMAIMOR 772
 QY 753 SDIFRVAIAGAPVTL 767
 Db 773 PQVFKAOPLAAPPRL 787
 RESULT 15
 US-09-976-674-37
 ; Sequence 37, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akusanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Junien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: US 60/240,117
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 37
 ; LENGTH: 819
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-976-674-37

Query Match 50.1%; Score 2356.5; DB 10; Length 819;
 Best Local Similarity 58.1%; Pred. No. 7.7e-217;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Inc.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:08:55 ; Search time 23.176 Seconds
(without alignments)
3659.853 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700
Sequence: 1 MAAMETERQLCVETFEFADDC.....HLHYLGENTGSRIALAKVI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922.5	19.6	931	2 T32919	hypothetical prote
2	842	17.9	738	2 A87516	dipeptidyl peptida
3	754	16.0	741	2 JC5142	X-Pro dipeptidyl-P
4	620	13.2	711	2 S66261	X-Pro dipeptidyl-P
5	529	11.3	766	1 CDH26	dipeptidyl-peptida
6	519.5	11.1	793	2 T41703	dipeptidyl aminope
7	518.5	11.0	760	1 S23752	dipeptidyl aminope
8	510.5	10.9	792	1 A30107	dipeptidyl-peptida
9	484.5	10.3	818	1 A30107	dipeptidyl-peptida
10	465	9.9	931	2 A49737	dipeptidyl aminope
11	451.5	9.6	799	2 T25174	hypothetical prote
12	445.5	9.5	779	2 T25173	hypothetical prote
13	440	9.4	803	2 A41793	dipeptidyl aminope
14	427	9.1	803	2 168600	dipeptidyl aminope
15	427	9.1	865	2 154331	dipeptidyl aminope
16	378	8.0	829	2 T19514	dipeptidyl aminope
17	374	8.0	795	2 F82858	hypothetical prote
18	374	8.0	795	2 T19514	hypothetical prote
19	331.5	7.1	759	2 T38593	fibroblast activat
20	293	6.2	743	2 T37700	probable dipeptidyl
21	282	6.0	742	2 C71137	hypothetical prote
22	269.5	5.7	631	2 H75007	probable acylamino
23	268	5.7	683	2 E87495	prolyl oligopeptid
24	255.5	5.4	709	2 B82580	alanyl dipeptidyl
25	249.5	5.3	657	2 E70025	hypothetical prote
26	236	5.0	622	2 F71174	hypothetical prote
27	223.5	4.8	632	2 E75057	peptidase (impor
28	217.5	4.6	536	2 E70289	acylaminoacyl-pept
29	200.5	4.3	591	2 H72474	probable acylamino
30	185	3.9	569	2 S74053	probable acylamino

30	183.5	3.9	608	2 F83397	probable peptidase
31	182.5	3.9	674	2 B84381	acylaminoacyl-pept
32	180.5	3.8	721	2 T09631	probable acylamino
33	180	3.8	667	2 A87711	prolyl oligopeptid
34	172	3.7	732	1 UC4655	acylaminoacyl-pept
35	171	3.6	732	1 S07624	acylaminoacyl-pept
36	170.5	3.6	572	2 F72455	probable acylamino
37	163.5	3.5	659	2 F72568	probable acylamino
38	161	3.4	676	2 C97775	acylamino-acid-rel
39	160.5	3.4	732	1 J01032	acylaminoacyl-pept
40	160.5	3.4	745	2 T33751	hypothetical prote
41	158	3.4	598	2 F84199	hypothetical prote
42	157	3.3	614	2 E75094	prolyl endopeptida
43	155.5	3.3	654	2 AD3183	peptidase (impor
44	155	3.3	629	2 T15945	hypothetical prote
45	154.5	3.3	828	2 G87564	hypothetical prote

ALIGNMENTS

Query Match	19.6%	Score 922.5;	DB 2;	Length 931;
Best Local Similarity	29.3%	Pred. No. 1.4e-57;		
Matches	282;	Conservative 145;	Mismatches 326;	Indels 211;
Gaps	41;			
QY	33	EPFVEXYSQSLKLLADIRKYGIMAKAPHDPMFVKRNDPPGHSRITYILAMSGEN	92	
DB	36	EPARFETRSFSQ--LIDHARSWTEVRGMTQGFTRISIMRAE--KDRIMYALISGVP	89	
QY	93	RENT--LFYSEIP-KTINRAAVLMLSWKPLDLFOATLIDVGMV-----SRE	135	
DB	90	GTNQSIFSVTIPELVEKQAVADRK-----FELTKSGYNVDSYIRMSCKRTPPSAE	142	
QY	136	EELIEREKRIQTV-GIASYDHHQ--SGTFLPFGA--	167	
DB	143	FTLQCEHRSQVVTGIDYEIRNGKMLMGDDQFRFNPLEALAIPIAVDDQSSTEP	202	
QY	168	-----SGTHVKRGQGG--FTQOPLRPMLVERSCP-----NIRMDPLCP	206	
DB	203	MDISEGSIITGKSCSNEAPQSSTVPVTRIPRIKPPFTTSKEPATAPTNVFSASAVCP	262	
QY	207	ADPDWIAFIHNSINDINISITVTRERRLTYVHNELANMEEDARSAGVTVLOEFDYSG	266	
DB	263	ADSSILAVLVNKQVYI-----EKNGKTIHRSSNSKHITN--GVPSTYVQELERPEG	313	
QY	267	YWCCKAETTPSGGKILRIIYENDESEVEIHHV-----TSPMLETRRADSFRRYKPTG	319	
DB	314	IWM-SESKT-----RLTYEHVNEKVAESQGVGVGDPVAPM-----KYPRAG	355	
QY	320	TANPKVTFKMSLEIMDAEGRIIDVIDKELIQPEILPEG--VEYIRAGWTPEGKAWS	376	
DB	356	TKNAYSTLRN---VILENGKAYDVPDKD-----EVIYKHCPEYETITRAAGFSDGTVMV	407	

```

QY 377 ILDRSOTRLQIVLI-----SPELFYVEDDVMERORLIESVP 414
D 408 QWMSDQAOCCILLIPYDPLPRELGSIKEDNIQSTDLNMGWMD-KSHEETMEKRP 466
QY 415 DSVT--PLIYBETDININHDIFHVPQSHHEE-IEFIPASECKTGFRLHYKTSIL 470
D 467 RGLKRGTVQIHKARDWYINTHNALYPLKITEDEHPMEFICYLEKPNG-SCLALISML 525
QY 471 KESKYKSSGGLPASPDKCPRIKEIATSGMEVLSGHSNIOQVDERRLVYEGTOS 530
D 526 DONG-----CRHTEKLLMAENFSI-NKSMGIIVDEVREILVYVAVNESH 569
QY 531 PLEHLLVYVYNPGEVTRLTDRGYSHSCISOHCDFISKYSNOK-----NPHCVSLYKL 586
D 570 PLEWMI-CVSHYRQGHQHLTESGI-----C-FKSRANGLALDLDHGFACVMT 617
QY 587 S--SPEDDPTCKYEF-W-----ATILDSAGP-LPD-YTPPELFSFES-TTG 627
D 618 SVGSPAE---CRFYSFRWKENEVLPSTVYANITVSGHPQDPDLHFDSPEMIEFQSKTG 674
QY 628 FTLYGMLYKPHDLOGKKYPTVLFYGGQVOIVNNRKYKVPRLNTLASLGYVYVID 687
D 675 LMHYAMILRPSNPDYKYPVPHVYGGPQIVHNDSTWQYLR---FCRLGYVYVVID 731
QY 688 NRGSCHRGKFEKGAFFKYMGGQIEIDQVEGLQYLASRY-DFIIDLRYGHSYGYSTL 746
D 732 NRGSANRGIEFERHIIKKMGTVVEVDQVEGLQMLAERTGPMDSRVVHHMSYGGTAL 791
QY 747 MALMORSIDIFRVALAGAPVTMTIYDTGYTERYMGHPDNOGYLGSVAMQAEKFPSEP 806
D 792 QMAKHPIYRAALAGAVSDMRLYDTAYTERYMGYP-LEEHVYGASSITGLVLEKLPDEP 850
QY 807 NRLLLLHGFDLENHFAHSTILSLFLVAGKPYDLQIYPOERHSIRPESGEHLEHLLH 866
D 851 NRMLVHGMNDENHFAHLEHVDCEIKKGMHLEVLPENRHGVRRNDASTIYLDARMY 910
QY 867 YLQGE 870
D 911 FAQCG 914

RESULT 2
A87516
dipeptidyl peptidase IV (imported) - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: A87516
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kjolseth,
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.;
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: A87516
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-738 <STO>
A/Cross-references: GB:AE005673; NID:913423647; PIDN:AAK24125.1; GSPDB:GN00148
C/Genetics:
A/Gene: CC2154

Query Match 17.9%; Score 842; DB 2; Length 738;
Best Local Similarity 29.9%; Pred. No. 5.5e-52;
Matches 221; Conservative 115; Mismatches 262; Indels 141; Gaps 24;

QY 133 SREBELRRKRIGTVGIAVDYHQSCTFLFQAGSGIY--HYVDGPGQGTQOPLRPNL 190
D 94 SEAEKARRBRARVASRGIYVSWDQGRFLIVPLDGLYDAVADGKITRLTLE----- 146

QY 191 VETSCPNIRMDPKLCPADPDMIAFIHNSDIWISNIVTREERRRLTYVHNELANEEDARSA 250
D 147 ----TGDEVDVAVKSPKG--GYVSVYRDNQNIYIKPVAGGALTALT-----TGKQALSF 194

```

```

QY 251 GVATFVLOEERDRSGYWCCKAETTSQCKILAILHEENDESEVEIHWTSPLMETRRA 310
D 195 GVAEFIVQEBELDRFTGYWMSDES-----RIVYTRVDESQVDIV-----PRA 236
QY 311 D-----SFRPKGTANPKVTFKMSIEMDAEGRI-IDV-IDELIQPEILLFEG 358
D 237 DIGFGATVNNQRIPPRAGRPAVVDLFVRDL--ASGVVALDLGANKD----- 283
QY 359 VEYIARAGWTPBGRKYSAILDRSOTRLQIVLISPELFIPEVDDVMERQRLIESVDSVT 418
D 284 --YVARVAMSADGKTIVYVQRSLRSDQKTLDLIAF-----DLAT 318
QY 419 PL-IYBETDININHDIFHVPQSHHEEIEFIPASECKTGFRLHYKTSILKESYK 476
D 319 GAGKTLITLDDPHFIEVSNFRPLTDG---FTLMGSE-KDGNQHLTRYA----- 363
QY 477 RSSGLPASPDKCPRIKEIATSGMEVLSGHSNIOQVDERRLVYEGKDSPLRHL 536
D 364 -ADGKTLA-----QITGDMFVIGLEG---VDEARRKVALFSASIDTPIERRL 406
QY 537 YVSVYNPGEVTRLTDRGYSHSCISOHCDFISKYSNOKNPHCVSLYKLSSP-----ED 591
D 407 YEVSYAKPRKPKALTSAGWMAAKVADNCGAFAGTSPDKTPSQGTALYSADGKRVWIEE 466
QY 592 DPTCKTEFPATILDSAGPLPDYTPPELFSFESTTGTLYGMLYKPHDLOGKKYPTVLF 651
D 467 NKLAEHGPYW---PYANLPQ---PEFGSKADAGETLHLEIKPFGFPAKKYPAIVS 519
QY 652 IYGGPQVOIVNNRKYKVPRLNTLASLGYVYVINDRSGCHRGKFEKGAFFKYMGGQIEI 711
D 520 YVGGPQVHQRWAKMHSSE---RFLYLAGYVIFKIDRSGNRSKAKFPARALDRKLGIVEV 576
QY 712 DDQVEGLQYLASRYDFIIDLRYGHSYGYSTLSMALMORSIDIFRVALAGAPVTWTFY 771
D 577 EDQLLGAKFPLASQ-PYVADADKLGYMGWSYGGFPMALMLLTAEPTFKGAGAGAPTEWGLY 635
QY 772 DTGYTERYMGHPDNOEGYVLSGAMQAEKFPSPRNLLIHGFDLENHFAHSTILSF 831
D 636 DTATYTERYMGHPDNRKAGVYSIDINRIDKL-APGSLLLHGMADNVIFENSTRMLAA 693
QY 832 LVRAGKPYDLQIYPOERHS 850
D 694 LQKRALIFEMAMYPGERHS 712

RESULT 3
JC5142
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C/Species: Xanthomonas maltophilia
C/Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C/Accession: JC5142
R:Kadashima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A/Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and expressio
A/Reference number: JC5142; MUID:97164011; PMID:9010758
A/Accession: JC5142
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-741 <RAB>
A/Cross-references: DDBJ:D83263; NID:91753196; PIDN:BA11872.1; PID:91753197
C/Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl re
C/Superfamily: dipeptidyl-peptidase IV
C/Keywords: dipeptidylpeptide hydrolase; membrane bound
F.4.18/Domain: transmembrane #status predicted <YMW>
F.610/Active site: Ser #status predicted
F.685/Active site: Asp #status predicted
F.717/Active site: His #status predicted

Query Match 16.0%; Score 754; DB 2; Length 741;
Best Local Similarity 28.4%; Pred. No. 1.1e-45;
Matches 214; Conservative 120; Mismatches 281; Indels 138; Gaps 26;

```

QY 133 SREELERKRGVIGT-GIASVDYHQSSTFLFOAGSG--IYHVKDGGQGTQOPLRPN 189
 Db 93 SDEEARRERORLIAMTGIYDOWSPDAGRLLFPLGELLYLDLQEGRAVROU----- 147
 QY 190 LVETSCPNIRMDPKLCPADPDWIAFIHSDIWIINSIVTEERRLTYVNHLEAMBEDARS 249
 Db 148 ---THGEPAITAKISPRG-GFVSFIRGNLWVIDLASRQWLT-----ADSGT 193
 QY 250 A--GVATFVLOEFEDRYSGYWPCRAETTPGGKILRLIYEBNDESEV-----EIIHV 300
 Db 194 TIGNIAEFVADDEEMDRHTGYWMAPDDSA-----IAYARIDESPVYQKREYVAD 244
 QY 301 TSPMETRRADSFRRPKXTGANTPKYTFKSEIIMABEGLIDVIDKELOPEILFEGVE 360
 Db 245 RTDVEQ-----RYPAGDANVQVKGIVISPAEQAQWIDLGKEODI----- 287
 QY 361 YIARAGWTPGKYASILLDRSOTRLQIVLISPFLFIPVEDDWMEROLISVPDVTPL 420
 Db 288 YLARWNRMDPOLHSQ-RQSRDQKLDLVEVLT-----ASNQR----- 325
 QY 421 IYETTDIWINIHDFHVPFQSHHEEIEFIFASCKTGFRHLKYKITSILKSKYKRSQ 480
 Db 326 VLAHETSPWVPLNHSRLFL-----DDGSIIMSSS-RFGQHLRYI-----DSKGA-- 372
 QY 481 GLPARDKCPKEIATSGEWEYLGRHGSNIQVDEVRLVYFEGTDSPLHHLHYVS 540
 Db 373 -----ALTHGNMSV-----DELLAVDERAGLAFPAIGESARBSQIYAVP 412
 QY 541 YVNGEVRTLTDGYSHSCISQHCDFPISKYNQKNPHCVSLYKLSPPEDPTKTEAF 600
 Db 413 -LOGGQFQRLSKAPMGHSASFARNASVYVDSMNSNSTPQJLEIFLANG-----EK 461
 QY 601 WATLL--DSAGLPPTY-----PELFSFESTTGF-LYGMLYKPHDLPQKKYPTVL 650
 Db 462 IATLVENDLADPKPYARFYARREARVEFEFTLTADGKTPLYNSVYKPAFGPAPKRYPAV 521
 QY 651 FIYGGPOVLYNNRFGVKYFRILNT-LASLGYVVVVIDNRGSCRHGLKFBGAFKYMQI 709
 Db 522 YVYGGPASQVYTDMPFGDHLFNQYLAQGGVIVFSLNRGTPRGRGFGALYKQGTIV 581
 QY 710 EIDDQVGLQYLASRYFDIDLRVGIHMSYGYLSLMAQMORSDFRVAIAGAPVTLMI 769
 Db 582 EVADQQLGVAWL-KQGPVWDPARIGVQGMNSGYTITMLLAKASDYSACGAGAPVDWG 640
 QY 770 FYDGYTFRYVGHDDQNEQGYLLGSVAMQAEKFPSEPNRLILLHGFLENVHAFTSL 829
 Db 641 LYDSHYTERYVDLPARRDAGYREARVLTIEGLRSP--LILLHGMADDNVLFTNSTSLM 697
 QY 830 SFLVAGKPYDLOIYPOERHSIRVPESGEHYEL 862
 Db 698 SALQKQGPFLMTYPPAKHGLSGADALHRYRV 730
 RESULT 4
 S66261
 X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
 C:Species: Flavobacterium meningosepticum
 C>Date: 28-Oct-1996 #sequence.revision 13-Mar-1997 #text-change 20-Jun-2000
 C:Accession: S66261
 R:Kabaashima, T.; Yoshida, T.; Ito, K.; Yoshimoto, T.
 Arch. Biochem. Biophys. 320, 123-128, 1995
 A:Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from *Flavobacterium meningosepticum*
 A:Reference number: S66261, MUID:55314307, PMID:7793970
 A:Accession: S66261
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-711 <KAB>
 A:Cross-references: EMBL:DA42121; NID:9577283; PIDN:BA07702.1; PID:9577284
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidylpeptide hydrolase
 Query Match 13.2%; Score 620; DB 2; Length 711;
 Best Local Similarity 25.2%; Pred. No. 4e-36;

Matches 195; Conservative 134; Mismatches 284; Indels 160; Gaps 26;
 QY 145 IGTGVIASVDYHQS-----SGTF-----LFOAGS-----GIYVXQG 176
 Db 54 IEPITIAKSYTSQKKNYIDGSGFTTNSDESKILLQSSQSIYRHSFLGKEVND 112
 QY 177 GPGQFTQOPLRP-NLVETSCPNIRMDPKLCPADPDWIAFIHSDIWIINSIVTEERRLTY 235
 Db 113 -----LKSRTVVSINNANWIOEPKFSF-DGSKVAFIADNNLFYQDLNKGTLQITL 162
 QY 236 --VNHLEAMBEDARSAGVATFVLOEFEDRYSGYWPCRAETTPGGKILRLIYEBNDES 293
 Db 163 DKKNEIILN-----GLGQWVVEEERGHADYQW-----NKAGDALVFRFDERKVP 208
 QY 294 EVEI-IHTS--PMLETRADSFRPKXTGANTPKYTFKSEIIMABEGLIDVIDKELO 350
 Db 209 EINITIYQNLVPLKLT-----YKPKKGEESAVATAYLYQ--SSG----- 249
 QY 351 PFEILFEGVE--YIARAGWTPGKYASILLDRSOTRLQIVLISPFLFIPVEDDWMERQ 408
 Db 250 SAQINFGSGSEKYYIDQLPQTWANDIIVATANRHQNKVDLKVNTK----- 295
 QY 409 LIESVPDVTPLIYEETTDIWINIHDFHVPFQSHHEEIEFIFASCKTGFRHLKYKTS 468
 Db 296 -----TAAVSKLFTETDNAMIEFTDNLTFE-----LDDNSFLWASR-RDGRHLWYDA 343
 QY 469 ILKSKYKRSQGLPAPSDFKPIKEIATISGEWEYLGRHGSNIQVDEVRLVYFEGTK 528
 Db 344 AGKLKK-----QVKGMEIINYYGNPKTE---VITQTE 377
 QY 529 DSPLEHLHYVS--YVNGEVRTLTDGYSHSCISQHCDFPISKYNQKNPHCVSLYK 586
 Db 378 KGSINK--VYSKININTGKTQLLSNABGNNSAFAFSKTFNFINSSIAKYPKILMDA 434
 QY 587 SSPE-----DDPTCKTKEWATILISAGLPDYTPPEITFSFESTTGF-LYGMLYKPHD 639
 Db 436 NGKQVYELQNNDDLNLKLS-----DNFAKEFTTIPPAADQWAMMIKPKN 482
 QY 640 LQPGKYPVLYFIYGGPOVLYNNRFGVKYFRILNTLASLGYVVVVIDNRGSCRHGLK 699
 Db 483 FDPAPKTFVPMFOYSGSGSQVANSWDGNGIMEDMLAQKGLLVVCVGRGFGYGYK 542
 QY 700 GAFYKKGQIEIDDQVGLQYLASRYFDIDLRVGIHMSYGYLSLMAQMORSDFRVA 759
 Db 543 KVTYKNLGKYEIEDQITAAKVLGNQ-SYVDSRIRIGIPMSYGGYNASLAKMGADVFMG 601
 QY 760 IAGAPVTLMTFYDGYTFRYVGHDDQNEQGYLLGSVAMQAEKFPSEPNRLILLHGFLEN 819
 Db 602 IAAVAVTNMRPYDSIYTERFLQTPQENKDGDLNSPTTYAKTL--KCKPILLHGTADDN 658
 QY 820 VHAFTSILLSFVAGKPYDLOIYPOERHSIRVPESGEHYELHLYLOENL 872
 Db 659 VHFQNSMFESEBALIQNKQDFPMAYPDKNHSIIQGNTRPQYKERTNYILENL 711
 RESULT 5
 CDH26
 dipeptidyl-peptidase IV (EC 3.4.14.5) - human
 N:Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)
 C:Species: Homo sapiens (man)
 C>Date: 31-Dec-1993 #sequence.revision 23-Aug-1996 #text-change 18-Jun-1999
 C:Accession: S24313; B42408; B42408; B61136; S59510; I56154; S59857; S15520
 R:Matsumi, Y.; Hayashi, Y.; Arakawa, F.; Ikehara, Y.
 Biochim. Biophys. Acta 1131, 333-336, 1992
 A:Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a serine protease
 A:Reference number: S24313, MUID:92329551, PMID:1352704
 A:Accession: S24313
 A:Molecule type: mRNA
 A:Residues: 1-6, '1', 8-766 <MIS>
 A:Cross-references: EMBL:X60708; NID:935335; PIDN:CAA3118.1; PID:935336
 J:Darmon, D.; Lacasa, M.; Batteault, L.; Marguet, D.; Sepin, C.; Troclet, P.; Barbé, A.
 J. Biol. Chem. 267, 4824-4833, 1992
 A:Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer


```

QY 373 YAMSTLL---DRSQRLOIVLISPEL--FIPVEDVWER-----ORLIESVDSVTPLELI 422
D 327 ---SVLVFTNRNSSTICFARLLDITKSIHTKTECEGWYEQSAAKMP--LNNSLV 381
QY 423 YEETDIDMINHIDIPHFPOSHEEIEFIPASECTGRRLHYKITSILKESYKSSGGL 482
D 382 WENMSD---GYDILALDDYNHAFIPF-----NSSS 410
QY 483 PAPSDKCPKEIEIAITSGEWEVLGRHGSNIQVDEVRRLVVEGTRKDSPLEHLLVYSYV 542
D 411 P-----IYLTSGAMDVT---DGIPIHDDPQNAVYFLATLKOSTERHLYYS-L 454
QY 543 NPEVTRLTDRGYS---HSCCISQCHDFPISKYSNQKNPHCVSLYKLSPEDDPTCKTKE 599
D 455 DTLEIYGLTDNGEDEGYSTSPGDFYVLYNHGPDVP---WQELRSTKDKOYCSLE 510
QY 600 FMATILDSAGPLPDVTPPEI---FSFESTTGFTLYGMLYKPHDQPKKYPVTLFTYGG 655
D 511 TNSRLKQO---LSSITLPSVEYKGLTFNDTT-FNF--MERRPRNDVNRKYPVLEFAYG 564
QY 656 PVOVLVNNRFKGVKYPRLNTLASLG-----YVVVVIDNRGSGHGLKFEGAFFYKMGQI 709
D 565 PGSQOV-----AKLFRVDFQAVLASHDPDFEIVVTLDRGTGFNGNAFRYSVRHLGEM 618
QY 710 EIDDOVEGLQYLAARVPIDLDRYGCHGMSYGYLSMALMQRSDIPRVAIAGAPVTLMT 769
D 619 ESYDQOQKGFMAW-LRPVDENHWGIMGWSYGYLTLEKTL-ETQDVFSYGAAVAPVTDWR 676
QY 770 FYDGTYERVYGHDPDQEGGYLGSVAMQAEKFPSEPRLLHLHGFDENHFAHTSILL 829
D 677 LYDSVTERVYMDLPQYNKEG-YKNSQIHDIYEF-KOLKRFVPAHGTGDVNHFGSHMLM 734
QY 830 SFLLVRAG-KPYDLQIYQERHSIRVPBSGEHYELHLHYLOENLSRIAL 879
D 735 DGLMLANCYNDMAVFPDSASHI---SYNNASLSIYHRSLEWIGDALGRI 781

```

RESULT 7
S23752
dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse
N/Alternate names: CD26 alpha subunit, THAM alpha subunit
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
R/Accession: S23752; A46465; A56030
R/Marguet, D.; Bernad, A.M.; Vivier, I.; Darmon, D.; Naquet, P.; Pierres, M.
J. Biol. Chem. 267, 2200-2208, 1992
A/Title: cDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-di-
A/Reference number: S23752; MUID:92129288; PMID:1370813
A/Accession: S23752
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-760 <MAR>
A/Cross-references: EMBL:X58384
R/Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernad, A.M.;
J. Immunol. 147, 447-454, 1991
A/Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase
A/Reference number: A46465; MUID:91302787; PMID:1712807
A/Accession: A46465
A/Status: preliminary
A/Molecule type: protein
A/Residues: 1-20 <VI>
A/Experimental source: M14.T thymoma cells, Swiss nu/nu
R/Bernad, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.
Biochemistry 33, 15204-15214, 1994
A/Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.
A/Reference number: A56030; MUID:95092780; PMID:7999781
A/Accession: A56030
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 746-760 <BER>
A/Cross-references: GB:U12620

C/Genetics:
A/Gene: CD26
C/Superfamily: dipeptidyl-peptidase IV
C/Keywords: dipeptidyl-peptidase hydrolase; glycoprotein; transmembrane protein
F.213,223,315,514,679/Binding site: carboxylate (asn) (covalent) #status predicted
F.624,702,734/Active site: Ser, Asp, His #status predicted

Query Match 11.0%; Score 518.5; DB 1; Length 760;
Best Local Similarity 25.0%; Pred. No. 8e-29;
Matches 203; Conservative 107; Mismatches 306; Indels 197; Gaps 37;

```

QY 131 MYSEBELRERKRIQTVGI-----ASYDHQSGTFLF-----OAGS 168
D 67 LYQENNLINAEHGNSSIFLENSTFSPGYHSVSPRLVLELVYVYKQMRHSYTSY 126
QY 169 GIHYKQGGPGGFQOPLRPLVETSCNIMDKLCPADWLT-----AFHSND 219
D 127 NIYVNR-----ROLTEKIPN-----NLOWTWSPEGHKLAVYWRND 165
QY 220 IWINIVTREBRRLTYVHNELANMEEDARSAGVATFVLOEE-FDRYSGVWNCPK----- 272
D 166 IYVVEPHLPBHRIT-----STGEENVITNGITDWYEEVFGAYSALMSPNNTFLAY 219
QY 273 AETTPSGKILRIIYEENDESEVEIIVTSPLETRRADSRFPKTYGTANPKYTPKMEI 332
D 220 AQFNDTGVPILIEYSF-YSDS-----LQPKYVMIYPRAGAVNPTVKFFI--V 265
QY 333 MIDAEGRIDVIDKELIQPEILFEGVEYIARAWTPBGKA-----WSILLDSQ 383
D 266 NIDLSSSSSAAPQIAPAPASA-RGDHYLDDVVAATEERISLOWRRIQWYSWAICDY 324
QY 384 TRLOIVLISPELFIPEDDVWERQRLIESVDSVTPRLIYEETDIDMINHIDIPHFVPS 443
D 325 DKIMLTNNCS-----EQQH-----EMSTGWVG--RFRPAEYHF 358
QY 444 HEEIEFTFASCTGRRLHYKITSILKESYKSSGGLPAPSPFK-CPIKEIAITSGE 502
D 359 TSDGSFFYKTIISDQGYKHCHF-----PKDKKQCTF-----ITRGA 395
QY 503 WEVLGRGSMNQVDEVARLYVFBGT-KDSPLEHLLVYVSYNPEVTRLTDRGYSHSCI 561
D 396 WEVL-----STEALTSPLYIISNQYKEMPEGRRLYKI-----QLTQ--HTNVKCL 439
QY 562 S-----CHCDFPISKYSNQKNPH---C-----VSLYKLSPEDDPTCKTKEFMATILDSAG 609
D 440 SCDLMPRCQYVAAVSFKAKAYQLGQMGRLPYTLHRSIDHEKLVLE-----DNSA 493
QY 610 ---PLPYTPPE---IFSFSTTGFTLYGMLYKPHDQPKKYPVTLFTYGGPQVQLVNN 663
D 494 LDRWLDQVQWPSKKLDPIVNETRF-WYQMIILPPIH-FDKSKKYPDLLDVVAGPSCQ---- 547
QY 664 RFGVKKYFRLN---TLASLGYYVVV-IDNRGSGHGLKFEBAFFYKMGQIFIIDQVEGLQ 719
D 548 ---KADASFRLNMAITYLASTNITIVASFDGSGSGQKIMHAINRRIGTLEVEDQIEAR 605
QY 720 YLASRYDFIDLDVGVHGMVSGYGLSLMALMQRSDIPRVAIAGAPVTLMTFYDGYTERY 779
D 606 QPV-KMGFVDSKVVVIMGWSYGYVSMVLSGSGGVKCGIIVAPVSRMEYDSVTERY 664
QY 780 MG--HFDQEGGYLGSVAMQAEKFPSEPRLLHLHGFDENHFAHTSILSLFVBRGK 837
D 665 MGLPIEDNIDHRNSTVMSRAHF--KQVEYLHHTADNVAHQFQSAISKVLVDAGV 722
QY 838 PYDLQIYQERHSIRVPBSGEHYELHLHYLOE 870
D 723 DFGAMVYTDDEHGIASSTAHQIYSHMSHPLQ 755

```

RESULT 8
A39914
dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat
N/Alternate names: GP110, membrane glycoprotein 110K; OK-61
N/Contains: dipeptidyl-peptidase IV, soluble form


```

QY 267 YWPCPAETTPSGKILRLIYEENDESEVEIITHVTSPLMETRAUS-FRYPKTGTANPKY 325
  |||
Db 362 IWM-----APDSSKAVFARFNDTSVDIRLNRYTN-MNBAYALPTDKIKYPKGFQNPQF 414
QY 326 TFRKSMIMDAEGRRIIDVDKELIQFELLFBEVEEIAAGWTPBEKRYAMSLIDRSQTR 385
  |||
Db 415 DL-----FLVNLQGIITYSINTG-GQKOSILYNG-----KWLSPDFFREI-TRNMSXI 461
QY 386 LQIVLISPELFTPVEDVNERQRLIESVDSVTPLLIYEETTDIWI-NHDIHFVPOSH 444
  |||
Db 462 LQVKYV-----DIPSSQML-TVNTNSNLF-----NGWLEKTKDLISIPKFE 503
QY 445 EEBIE-FTIPASECTGFPHLYKITSILKSKYKSSGGLPAPSDCKPIKEEIAITSGE 502
  |||
Db 504 LKRMQGYIDIHADSRGFSHLFYFPTVF-----AKEPDQITKN 542
QY 503 WEVLGQH--GSNIQVDEVKRLVFEECTKOSPLEHLYVYSV-----NPEVYT 548
  |||
Db 543 WBTGTGIVGYEYETD---TIFFTANEIGVMSQHLYSISLTDSTQNTQFOSLQNP--- 594
QY 549 RLDRGYSHSCCISOHCDFFISKYSNQKNP-----HCVSLYKLSPEPD 592
  |||
Db 595 --SDKXDFDYDFELSSARAYAIKSLGPDTPIKVAGPLTVANVAEIHDSIIQLTYDE-- 650
QY 593 PLCKTKEFWATILDSAGPLPDYTPPEIFSESTT--GFTLYGMLYKPHILOPKKXPTV 649
  |||
Db 651 ---KFKRE-----KIKNYDLR-ITSYKTMVLDGVEINVIYIRIKPAINLPKKKYPTL 696
QY 650 LFIYGGPQOVLVNNRPFKGVKYFRLNTLASLGYVVVVIDNRGSGHGLKEGAFYKMGQI 709
  |||
Db 697 VNIYGGPQSQTFTT--KSLAEQAVVSGLDIVIQIEPRGTGKGKWSRSRARELGTW 754
QY 710 EIDDOVE-GLQYLASRYDFIDLRVGIHGSYGYLSMAL-MORSDFRVAIAGAPVTL 767
  |||
Db 755 EPRDITEVTKKFIQNSQIIDSXIALIMWSYGGFSLKTVBLNDGDTFKYMAAVALPTN 814
QY 768 WIFDTGTERKMGHPDQNEQGYIYGSVAMQAEKPPSENNRLLILHGLDENVHFAHTSI 827
  |||
Db 815 WLYLSYVTERYMNQPSSEHNGEYFVSTIQNFKSES-LKRLLFIHGFDDVHIIQNTFR 873
QY 828 LLSFLVRAG-KPYDQIYPOEHSIRVPESGHEYLHLHYQENLGSIALAKXI 882
  |||
Db 874 LVQDLNLGLTNYDWHITFPDSHSIRYHNAQRIYQKLYLMDPAERFNTFVL 929

RESULT 11
T25174
hypothetical protein T23F1.7b - Caenorhabditis elegans
C1Species: Caenorhabditis elegans
C1Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
R1Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A1Reference number: Z19990
A1Accession: T25174
A1Status: preliminary; translated from GB/EMBL/DBJ
A1Molecule type: DNA
A1Residues: 1-799 <Wtl>
A1Cross-references: EMBL:Z81129; P1DN:CA803412.1; GSPDB:GN00023; CESP:T23F1.7b
A1Experimental source: clone T23F1
C1Genetics:
A1Gene: CESP:T23F1.7b
A1Map position: 5
A1Introns: 13/3; 52/1; 111/3; 197/3; 366/3; 524/1; 557/1; 648/2; 703/1; 743/1
C1Superfamily: dipeptidyl-peptidase IV

Query Match          9.6%; Score 451.5; DB 2; Length 799;
Best Local Similarity 23.0%; Pred. No. 5.4e-24;
Matches 195; Conservative 129; Mismatches 298; Indels 227; Gaps 38;

QY 78 PHSRPIYLAWS-----GENRENTLYSEIPKTIINRAAVLMSKELL-----DLF 123
  |||

```

```

Db 116 PSADKKYFAMDMHADNPQKNPQNETFHLKIYNNNERL-----NPLLFEVEELLFREL 168
QY 124 QATLDYGMWSREBELIREKRICTGYGIASYVDHQSGFTLFOAGSGIYHVKGQGFQ 183
  |||
Db 169 DSRTIYDGLRKESV-----IQAFKNGKFNDFVESNKIY----- 207
QY 184 QPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSDNIWISNIVTREERLLTYVHELAM 243
  |||
Db 208 -----QSSP-----EEBLTVNSNGEHT 226
QY 244 EBDASAGYATFVLOE-FDRYSGYWCKPAAETTPSGKILRIYEENDESEVEIITHVTS 302
  |||
Db 227 VD-----GLFDWITYEEEIFGRKDAWMSGK-----GQLAYASYDNHLTKVSL--KTY 273
QY 303 PMLETRRAD-SFRYEKT-----GTANPKVTFPKSMIMDAEGRRIIDVDKELIQFELL- 355
  |||
Db 274 HRLBEPYPIQTNPHYKTPAKVLPYTLISWNKTB-----QSRQDLVOLKDSLHYHLLA 328
  |||
QY 356 -----FEGVEYIARAGWTPBEKRYAMSLIL-----DRSQRLQIVLISPELFTPVEDVME 405
  |||
Db 329 VKMLINGTBOVSV-WT--NRYQNEVALTICDMQTAICRLEF----- 369
QY 406 RQRLIESVDSVTPLLIYEETTDIWINIHDFHVPQSHHEEIEFTIPASECTGFPHLYK 465
  |||
Db 370 -----KYASKRWVT-HDFHSHI-TSFEDTLFELLP-----HDKR 402
QY 466 ITSILKSKYKSSGGLPAPSDCKPIKEEIAITSGEMVEYGRHSGNIOVDEVRLRYE 525
  |||
Db 403 DNAFQVQVSLRSHQQLTPK-----FLNAGEYDVYSINGINKET---RTIFH 448
QY 526 GTKSPLEHLYVYSVYVNPGEVTRLDRGYSHSCIS---QHCDFFISKYSNQKNPHCVS 582
  |||
Db 449 AAAPKPSHSLFSYS-----LADSRNSAYCISCSINKCTMAQOMDMQMTAIVS 499
QY 583 LYKSSPEPD-----DPTCKT-----EFWATILDSAGPLPDYTPPEIFSESTTG 627
  |||
Db 500 CKGPAAPHTAIYNLTRMDSDKTEHANLLYKTYQNRVEAG-LPVIIRKETI--KISDD 555
QY 628 FTLVGMLYKPHILOPKKXY---PTVLFIYGGPQOVLVNNRPFKGVKYFRLNTLASLGYVV 684
  |||
Db 556 FDALIKLIPKDIYKRDHQALPLIVHYGGND--NTEATYIGIEVVAASQAAT 612
QY 685 V-IDNRGSGHGLKEGAFYKMGQIEIDDOVEGLQYLASRY-DPIDLRVGIHGSYGG 742
  |||
Db 613 LRIDRGSGGGRWKRSAIYQGLTVEVEDQIKAIKVILRLYRHLLDARVAVFQWSYGG 672
QY 743 YLSIMALMQRSD-IFRVAIAGAPVTLFIYDTGTERKMGHPDQNEQGYIYGSVAMQAE 801
  |||
Db 673 FMTLSMVNEAPQEFKCAVAPVTFAYYATYTERYWG--DAPLESY--SDVTKKLDN 728
QY 802 FPSENNRLLILHGLDENVHFAHTSILSLFVRAGKPYDQIYPOEHSIRVPESGEHYE 861
  |||
Db 729 FKS--TRLLHGHGLDDNVHFNQSAIILDELQNKGVDPDLNMYFQNASLSRTS--HYV 784
QY 862 LHLHYLOE 870
  |||
Db 785 GKMTHTFLRQ 793

RESULT 12
T25173
hypothetical protein T23F1.7a - Caenorhabditis elegans
C1Species: Caenorhabditis elegans
C1Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
R1Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A1Reference number: Z19990
A1Accession: T25173
A1Status: preliminary; translated from GB/EMBL/DBJ
A1Molecule type: DNA
A1Residues: 1-779 <Wtl>
A1Cross-references: EMBL:Z81129; P1DN:CA803411.1; GSPDB:GN00023; CESP:T23F1.7a

```

A:Experimental source: clone T23F1
C:Gene: CESP:T23F1.7a
A:Map position: 5
A:Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1
C:Superfamily: dipeptidyl-peptidase IV

Query Match 9.5%; Score 445.5; DB 2; Length 779;
Best Local Similarity 23.0%; Pred. No. 1.4e-23;
Matches 193; Conservative 126; Mismatches 292; Indels 229; Gaps 38;

```
QY 78 PHSRIYIAMS-----GENRENTLFYSEIPKTIINRAVLM,SMKFLDLFOATLDYMY 132
DB 116 PSARKRYAMMDHAPNPGMNFQNETFLKTIANNR-----ITYDGL- 158
QY 133 SREBLLERKRIGTVAGSYDHOQSGTFLPQAGSGIYHKDGGPGFTQOPLRPLVE 192
DB 159 -RKESV-----IQAFKMGKFNDFVFNKIKIY----- 187
QY 193 TSCNINRMDPLCPADPWIMAFHNSNDIWSINIVRERRLTLYVHNEIANNEEDARSQV 252
DB 188 QSSP-----EEGGLTRVSNNGEHTVD-----GL 210
QY 253 ATFLVLOE-PDRYSGYWMCPKAETTPSGKILRLIYENDESEVEIIVTSPMLETRAD 311
DB 211 FDMYIEEIFGRKXAMWMTK-----GDQAYASYDNLTKVSL-KTYHRLPPYID 262
QY 312 -SEFYPT-----GTANKVTFKSEIMIDABGRIDVIDKELIQPEIL-----FEGV 359
DB 263 TNFHYPTFAKVLPTLYLISNKKTE-----QSRQLDQOLDSLYHLLAVKMLEINGT 317
QY 360 EYIRAGWTPKGAWSIL-----DRSQTRLQIVLISPELFIPEDDVMERQRLIEVP 414
DB 318 EQLVSV-WT-NRQNEVALTICMDIATCRLEFE----- 349
QY 415 DSVPLIYEETDITMINIHDIHFVFPQSHHEIEFIIPASECKTGFRHLYKITSILKSK 474
DB 350 -----YKXASKRWLT-HDDFHSI-TSFEDTLFLLP-----HDKDNMFCQVAS 391
QY 475 YKRSSGGLPAPSDKCPKEIKIATISGMEVLAGRHGNSIQDEVRLRYFEGTQDSPLH 534
DB 392 LRLSHGQLRTEK-----PLNLGHYDVTISINGINKET-----RTIFFHAAPKPSHR 437
QY 535 HLIVYVSVNPEVTRLDRGYSHSCIS---OHGDFEFSKXSNQKNPCVSLYKLSGED 591
DB 438 SLFSYS-----LADSRNSAYCISCSLKNCTMQADQMDQKTAIVSCCKPAABHT 488
QY 592 -----DPTCKTK-----EFWATILDSAGPLPYTPPEIIFSFEFTGFTLYGMLYK 636
DB 489 AIVNLTMDSDKTEHANLIDYKTYQNRVEAG-LPVIKETI---KISDFDLILKSI 544
QY 637 PHDLPQPKKY---PTVLFYGGPOVLVNNRFKGVKTFRLNTLASLGVVVV-IDNRGSC 692
DB 545 PKDLYNDKQIALPLIVHYVGGPNDQ---NTEKATQIGIEBVVASASQALILRIDGSG 601
QY 693 HRGLKFGAFKYGKQIETIDQVEGLQYLASRY-DFTDLDRVGIHMSYGGYLSLMLMQ 751
DB 602 GRGKRYSAIYGOGLYVEVDQIKAIKIVLRLYHLLDARVAVFGMSYGGFMTLSMNE 661
QY 752 RSD-IFRVALIAGAVTLMIIFYDTGYTERYMGHPDQNGYGLSVAMQAEKFPSEPNRL 810
DB 662 APEOFFKCAVSAVAVTFMAYADATYTERYMG--DAPLESY--SDVTYKLDNFKS--TRL 715
QY 811 LILHGLFDENVHFAHTSILSLFLVRAKGPYDQIYPOERHSIRVPSEGEHYELHLIHLQ 870
DB 716 LMGHGLDDNVHFOHSAIILDELQNRGVDFLMLVYPNOAHSLSRTS--HVVGKMTFLRQ 773
```

RESULT 13
A41793
dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: A41793
R:Wada, K.; Yokotani, N.; Hunter, C.; Doi, K.; Wenthold, R.J.; Shimaaki, S.
Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992
A:Title: Differential expression of two distinct forms of mRNA encoding members of a di-
A:Reference number: A41793; MUID:92108018; PMID:1729689
A:Accession: A41793
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-803 <MAD>
A:Cross-references: GB:M76429; NID:9408719; PIDN:AA041623.1; PID:9408720
A:Note: sequence extracted from NCBI backbone (NCBI:75138)
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase; glycoprotein
F:257/342/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 9.4%; Score 440; DB 2; Length 803;
Best Local Similarity 24.4%; Pred. No. 3.6e-23;
Matches 176; Conservative 114; Mismatches 317; Indels 114; Gaps 29;

```
QY 171 YHVKDGGPQFTQOPLRPNLVETSCPNIRMDPKLCPADPWIMAFHNSNDIWSINIVTRE 230
DB 156 YVVLISKIPHGDPQSLDPEEVNNAKIQVAGWPK-----GOOLIFENNIIYCAHVQKA 210
QY 231 RRLTYVHNEIANNEEDARSAGATFVLOE-FDRYSGYWMCPKAETTPSGKILRLIYEE 289
DB 211 IRV-----VSTGKGVYVNGSDMLYEEILKTIAMWSPDG-----TRLAYAT 255
QY 290 NDESEVELIHV---TSPMLETRADSPRYPTGTANPKYTFKSEIMIDABGRIDVIDK 346
DB 256 INDSRVPMELPTGTSGYPT-AKPHYPAAGENPISLH---VIGLNGPHTDL--- 306
QY 347 ELIOPPEILEFGEVYIIRAGWTPKGYA-WSILDSQTRLQIVLISPELFIPEDDVM 404
DB 307 EMTPDDPDMREX-YITVWKWATSTKYAVNV--LSBAQN-----VS----- 344
QY 405 ERQLISVDPSPVPLII--YEETDITMINIHDIHFVFPQSHHEIEFIIPASECKTGRH 462
DB 345 -----LTLCDATGTCCKHEDSEAWL-----HRQNEEVFS--KDG-RK 383
QY 463 LYKITSILK--ESKYKSSGGLPAPSDKCPKEIKIATISGMEVLAGRHGNSIQDEVRL 520
DB 384 FFFRALPQGGQGFHYHITYSSSQPNSSNDIQ---SITSGDWVT---KILSYDEKRS 436
QY 521 LVYEGTKDSPLEHLLIVSVNDEVTR-----LDRGYSHSCISQHCDFISKYSN 574
DB 437 QIYFLSTEDLPRRQLYSASTV--GSFNRQCLSDLYNCTYFSAFSPGADFLKCEG 494
QY 575 QKNHCYSLYKLSDEDDPCKTKE-FWATILDSAGPLPYTPPEIIFSFEFTGFTLYGM 633
DB 495 PGVP-TVSVHNTYDKKMPDLEINEHVOKAISDRQMPKVEYRKLIE-----TDDYNLP 547
QY 634 LYKEDLPQPKKYPTVLFYGGPOVLVNNRFKGVKTFRLNTLASLGVVVV-IDNRGSC 693
DB 548 ILKATPTDPAHYVLLLVVGTGTSQVAKEF-AVMEWTVMVSHGAVVYKCGRGSGF 605
QY 694 RGLKFBFAFKYKQIETIDQVEGLQYLASRYDFIDLVYGIHMSYGGYLSLMLMQRS 753
DB 606 QGTRLAHEVRRRGLSELEKQMEAVRVML-KEPYIDKTRVAVFGKDYGYLSTYLLPAKG 664
QY 754 D-----IFRVALIAGAVTLMIIFYDTGYTERYMGHPDQNGYGLSVAMQAEKFPSEPNRL 809
DB 665 DGAQPAVSCGASALPTIDFKLYIASERELGLHGLNRAVEAKVHRSAL--EGQOF 722
QY 810 LILHGLFDENVHFAHTSILSLFLVRAKGPYDQIYPOERHSIRVPSEGEHYELHLIHLQ 869
DB 723 LVHATADKIHFOHTAELLTQILKGRANSLQIYPPESHYFSSAALQHLHRSILGFV 782
QY 870 E 870
DB 783 E 783
```

RESULT 14

168600
 dipeptidyl aminopeptidase like protein - human
 C1:Species: Homo sapiens (man)
 C1:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
 C1:Accession: 168600
 R.Yokotani, N.; Doi, K.; Wenthhold, R.J.; Wada, K.
 Hum. Mol. Genet. 2, 1037-1039, 1993
 A1:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat
 A1:Reference number: 154331; MUID:93372805; PMID:8103397
 A1:Accession: 168600
 A1:Status: preliminary; translated from GB/EMBL/DBJ
 A1:Molecule type: mRNA
 A1:Residues: 1-803 <RES>
 A1:Cross-references: GB:M6860; NID:9306707; PIDN:AAA3761.1; PID:9306708
 C1:Superfamily: dipeptidyl-peptidase IV

Query Match 9.1%; Score 427; DB 2; Length 803;

Best Local Similarity 23.4%; Pred. No. 3.1e-22;

Matches 172; Conservative 121; Mismatches 299; Indels 144; Gaps 30;

171 YHVKGDPQGTQPLRNVLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTSEE 230
 156 YVLSKIPHPDQSLDPEVSNAKLQYAGWGPX-----GQQLIFIFENNIYCAHVGRQA 210
 221 RLTYVHNEILAMMEDASAGATVLOEE-FDRISGYWMCCKAETTPSGKILRIIYEE 289
 211 IRV-----VSTGKGVLYNGLSDWLYEELIKTHIAMWSPDG-----TLAYAA 255
 290 NDESEVEIIHV---TSPMLETRRADSPRYPKGTANPKYTFMSEIMI DAGRIIDVTDK 346
 256 INDSRVPIPELPTVGTSTYPT--VKPHYPKAGSENPSISLH---VIGLNGPHTDL--- 306
 347 ELIQPEILFEGVEYIARAGWTPGKVA--WSILDRSQTRLOVLISPELFIPEVDVM 404
 307 EMPPDDPRMREY-YITWVKATSTKVAVTW--LNRAQN-----VS----- 344
 405 ERQRLIEVSPDVPPLII--YEETDWINIHDIHVAPQSHBEI-----EPIFASEC 456
 345 -----ILTLCATGTGCTCKKHEDESEAWLH-----RONEBVPFSDKGRKFFIRAI 390
 457 KTGRF-HLYKITSILKESKYRSSGGLPAPSDFKPIKEEIAITSGEWEVIGRHGNSIQV 515
 391 POGGRGKFYHIT--VSSSQPNSSNNIQ-----SITSGMDVT---KILAY 431
 516 DEVRILVYFEGTKDPSLEHNLVYVSYNPGEVTRLTDRGYSH--SCCISQHCDEFISKY 572
 432 DEGNKTIYFLSTEDLPRRRLYSAN-----TEGNFNRQCLSCDLVENCTYFSASF 481
 573 SNQKNPHCVSLYKLSPP-----EDDPTCKTKEF-----WATILDSAGPLDPYTPPE 618
 482 SHSMD---FLLKCEGPGVPMVTVHNTTDKKKMFLETHNEHVKKAINDRQMPKVEYRDIE 538
 619 IFSEESTTGFLLYGMLYKPHDLOPGKKYPTVLFYIGGPQVQLVNNRFPKGVYFRLNTLAS 678
 539 I-----DDYNLPMQILKPATFTDTTHYPLLLVVDGTGSGSVAKEFE--VSMWEVWVSS 590
 679 LGYVVVVIDNRGSGCHRGKKEFGAKFYKMGQIEIDDOVGLQYLASRYDFILDRVGHGW 738
 591 HGAVVVKCDGRSGSQGTKLHBEVRRLGLEEKDQMEAVRTML--KEQYIDRTVAVAFGK 649
 739 SYGGYISLMALMORSD---IFRVALAGAPVTLMIFYDTGYTERYMGHPDQNEGGYILGS 794
 650 DYGGYLSYTLIPAKGENOGQFTTGSALSPITDFKLVAASFSEKYLIGHGIDNRAAYEMTK 709
 795 VAMQAEKFPSPBNRLLLHGFLENDVNFATSIISLFLVRAKGYDQIYIPEQSHSIRVP 854
 710 VAHRVVAL--EEQGFLLIHPTADEKIHFOHTABELITOLIRKANYSLQIYVDESHYFTSS 767
 855 ESGEYELHLHYLOE 870
 768 SLKQHLVRSITNPFVE 783

RESULT 15

154331

dipeptidyl aminopeptidase like protein - human

C1:Species: Homo sapiens (man)

C1:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000

C1:Accession: 154331

R.Yokotani, N.; Doi, K.; Wenthhold, R.J.; Wada, K.

Hum. Mol. Genet. 2, 1037-1039, 1993

A1:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-rela

A1:Reference number: 154331; MUID:93372805; PMID:8103397

A1:Accession: 154331

A1:Status: preliminary; translated from GB/EMBL/DBJ

A1:Molecule type: mRNA

A1:Residues: 1-865 <RES>

A1:Cross-references: GB:M68659; NID:9306705; PIDN:AAA3760.1; PID:9306706

C1:Superfamily: dipeptidyl-peptidase IV

Query Match 9.1%; Score 427; DB 2; Length 865;

Best Local Similarity 23.4%; Pred. No. 3.4e-22;

Matches 172; Conservative 121; Mismatches 299; Indels 144; Gaps 30;

171 YHVKGDPQGTQPLRNVLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWSNIVTSEE 230
 218 YVLSKIPHPDQSLDPEVSNAKLQYAGWGPX-----GQQLIFIFENNIYCAHVGRQA 272
 221 RLTYVHNEILAMMEDASAGATVLOEE-FDRISGYWMCCKAETTPSGKILRIIYEE 289
 273 IRV-----VSTGKGVLYNGLSDWLYEELIKTHIAMWSPDG-----TLAYAA 317
 290 NDESEVEIIHV---TSPMLETRRADSPRYPKGTANPKYTFMSEIMI DAGRIIDVTDK 346
 318 INDSRVPIPELPTVGTSTYPT--VKPHYPKAGSENPSISLH---VIGLNGPHTDL--- 368
 347 ELIQPEILFEGVEYIARAGWTPGKVA--WSILDRSQTRLOVLISPELFIPEVDVM 404
 369 EMPPDDPRMREY-YITWVKATSTKVAVTW--LNRAQN-----VS----- 406
 405 ERQRLIEVSPDVPPLII--YEETDWINIHDIHVAPQSHBEI-----EPIFASEC 456
 407 -----ILTLCATGTGCTCKKHEDESEAWLH-----RONEBVPFSDKGRKFFIRAI 452
 457 KTGRF-HLYKITSILKESKYRSSGGLPAPSDFKPIKEEIAITSGEWEVIGRHGNSIQV 515
 453 POGGRGKFYHIT--VSSSQPNSSNNIQ-----SITSGMDVT---KILAY 493
 516 DEVRILVYFEGTKDPSLEHNLVYVSYNPGEVTRLTDRGYSH--SCCISQHCDEFISKY 572
 494 DEGNKTIYFLSTEDLPRRRLYSAN-----TEGNFNRQCLSCDLVENCTYFSASF 543
 573 SNQKNPHCVSLYKLSPP-----EDDPTCKTKEF-----WATILDSAGPLDPYTPPE 618
 544 SHSMD---FLLKCEGPGVPMVTVHNTTDKKKMFLETHNEHVKKAINDRQMPKVEYRDIE 600
 619 IFSEESTTGFLLYGMLYKPHDLOPGKKYPTVLFYIGGPQVQLVNNRFPKGVYFRLNTLAS 678
 601 I-----DDYNLPMQILKPATFTDTTHYPLLLVVDGTGSGSVAKEFE--VSMWEVWVSS 652
 679 LGYVVVVIDNRGSGCHRGKKEFGAKFYKMGQIEIDDOVGLQYLASRYDFILDRVGHGW 738
 653 HGAVVVKCDGRSGSQGTKLHBEVRRLGLEEKDQMEAVRTML--KEQYIDRTVAVAFGK 711
 739 SYGGYISLMALMORSD---IFRVALAGAPVTLMIFYDTGYTERYMGHPDQNEGGYILGS 794
 712 DYGGYLSYTLIPAKGENOGQFTTGSALSPITDFKLVAASFSEKYLIGHGIDNRAAYEMTK 771
 795 VAMQAEKFPSPBNRLLLHGFLENDVNFATSIISLFLVRAKGYDQIYIPEQSHSIRVP 854
 770 VAHRVVAL--EEQGFLLIHPTADEKIHFOHTABELITOLIRKANYSLQIYVDESHYFTSS 829
 855 ESGEYELHLHYLOE 870
 830 SLKQHLVRSITNPFVE 845

Search completed: October 15, 2003, 17:14:45
Job time : 31.176 secs

0 *

0 *

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 ; Search time 13.9931 Seconds
(without alignments)
2964.152 Million cell updates/sec

Title: US-10-070-464-1
Perfect score: 4700

Sequence: 1 MAAAMTEBOLGVEIFETADC.....HLHYIQENIGSRIMALKYI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	569.5	12.1	765	1	Q9N217 felis silve
2	566	12.0	765	1	P81425 bos taurus
3	525	11.2	766	1	P27487 homo sapien
4	523.5	11.1	767	1	DP4_RAT
5	519.5	11.1	793	1	Y8A8_SCHPO
6	517.5	11.0	760	1	DP4_MOUSE
7	484.5	10.3	818	1	DAP2_YEAST
8	478.5	10.2	761	1	SEPR_MOUSE
9	465	9.9	931	1	ST13_YEAST
10	454	9.7	760	1	SEPR_HUMAN
11	440	9.4	863	1	DP6C_BOVIN
12	435	9.3	804	1	DP6C_MOUSE
13	431	9.2	859	1	DP6C_RAT
14	427	9.1	865	1	DP6C_HUMAN
15	293	6.2	853	1	YDZF_SCHPO
16	249.5	5.3	657	1	YUXL_BACSU
17	172	3.7	732	1	ACPR_HUMAN
18	171	3.6	732	1	ACPR_RAT
19	161	3.4	732	1	ACPR_PIG
20	150	3.2	740	1	YL31_CAEEL
21	139	3.0	689	1	PPCE_AERHY
22	134	2.9	722	1	Y174_RICPR
23	134	2.9	1014	1	BGAL_BACDR
24	125.5	2.7	436	1	PAPA_CAVPO
25	124.5	2.6	1835	1	DURI_YEAST
26	123.5	2.6	705	1	PPCF_FLAME
27	123	2.6	1702	1	DPOI_THIELI
28	122	2.6	705	1	Y4S0_RHISN
29	119.5	2.5	1019	1	CA26_HUMAN
30	118.5	2.5	705	1	PPCE_FLAME
31	118.5	2.5	710	1	PPCE_HUMAN
32	118.5	2.5	1436	1	DPO3_STEAM
33	118.5	2.5	1438	1	DPO3_STEAM

34	118	2.5	1265	1	TOXA_PASMU
35	115	2.4	1047	1	POL_STVAL
36	115	2.4	3144	1	VP13_YEAST
37	114.5	2.4	710	1	PPCE_MOUSE
38	113.5	2.4	2483	1	COA2_HUMAN
39	112.5	2.4	835	1	NLG2_HUMAN
40	111.5	2.4	465	1	EGLC_RHIME
41	110.5	2.4	836	1	NLG2_RAT
42	110.5	2.4	4349	1	FAT2_HUMAN
43	110	2.3	1295	1	BXAI_CLOBO
44	110	2.3	1861	1	MAP2_RAT
45	110	2.3	2410	1	MOXI_SCHPO

ALIGNMENTS

RESULT 1
ID DP4_FELCA STANDARD; PRT; 765 AA.
AC Q9N217;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26).
GN DP4 OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripherel blood;
RA MEDLINE=20094000; Pubmed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA Mikami T., Takahashi E.;
RT Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT activation antigen CD26 homologue.";
RL Immunogenetics 50:366-368 (1999).
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline (By similarity).
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-
CC xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -1- PTM: The soluble form (SDP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AB023952; BAA92344.1; -
CC MEROPS: S09.003; -
CC InterPro: IPR002469; DPPIV N term.
CC InterPro: IPR001375; Peptidase S9.
CC InterPro: IPR002471; Pro_endopep_ser.
CC InterPro: IPR000379; Ser_estrs_site.
CC Pfam: PF00930; DPPIV_N term; 1.
CC Pfam: PF00326; Peptidase S9; 1.
CC PROSITE: PS00708; PRO-ENDOPEP_S9; 1.
CC Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
CC Transmembrane; Glycoprotein; Signal-anchor.
CC CHAIN 1 765
CC DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM

```

FT CHAIN 38 765 (MDP).
FT DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT (SDP) (BY SIMILARITY).
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 30 765 EXTRACELLULAR (POTENTIAL).
FT ACT SITE 629 629 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 765 AA; 88213 MW; 3EFCB98A22B175D9 CRC64;

Query Match 12.1%; Score 569.5; DB 1; Length 765;
Seed Local Similarity 23.7%; Pred. No. 8,2e-34;
Matches 225; Conservative 114; Mismatches 316; Indels 221; Gaps 42;

71 KRNDPDGPHSDRIYVLAAMSGENRENTLFYSEIPKTNINBAAVMLSMKPLDLFOATLDYG 130
30 KGNDA-AADSRRTYTLDTYLNKTRFKVYS-----LFM-----VSDHDY- 67
131 MYSREELLRRKRIGTVGI-----ASYDHOGS-----GTEF-----F 164
68 LYKQNNMILLFNAEYNGSSIFLENSTPDEFESINDSVSPGQITILEYVVKQMRHSY 127
165 QAGSGITHVKGQPGFTQOQRLPVLTSCNITMDKLCPADDWI-----AFI 215
128 TASYDIYDLNK-----RQITTEKIPN-----NQWTWSPBEGHKLAVY 166
216 HSNDIWISNIATREERRLTYVHNELANMEEDASAGVATFVLOEE-PRYSGYWMCPRAE 274
167 WKNDYVYKNEPNSSSHRTWTG-----EENAIYNGIADWYEEIIFAYISALMWSPGCT 220
275 TTPSGGKILRIIYENDESEVEIIT--VTSBMLETRRADSERIRPKXTANPKVTFKXSE 331
221 F-----LATYAFNDTQVPLIEFSYSDESIQYPTMTIRPYKAGANAP--TVKLFV 269
332 IMIDAEGRIIDVIDKEILIQFEILPEGVEYIARAGMTEGKYAMSILLDRQTRLQIVLI 391
270 IKTDNLNNTNATSVETTPPAAML-TGDIYLCQVYMANEERIS----- 311
392 SPELFIVEDDVMERQRIIESVDSVTPLIYEETTDIMINIHDFHVPQSHHELEFI 451
312 -----LQWLRRIQWY--SYMDIRDYNNGTGMIS-----SAAQHIMM- 347
452 FASECKTGF-----RLHYKITSILKESKYKSSGGGLPAPSPFKPIKEI 496
348 -----STTGWGRFRPAEPHFTSDGKNFYKIS--NEDGYKHI-----CRFQIDKDKCT 394
497 AITSGEMEVLRGHSNIQVDEVRLVYPEGT-KDSPLHHLHYVVSYVNPGEVTRLTDGTY 555
395 FITGAMWEVIG-----IEALTDTYIYISNEYKMGPGRNILYKI-----QLND--Y 438
556 SHSCGIS-----QHCDFFISKYSNOKNPH--C-----VSLYKLSPEDDPTCKTKFEWAT 603
439 TKVACTSCELKPEKCOYVSFSKAKYQURCGPGPLTYLTRSSNDELRVLED-NS 497
604 ILDASGRLPDYTPRE--IFSFESTTGFTLGMLYKHPDLQPKKKYPLVFLIYGGPQVQL 660
498 ALDKM--LQEVQMSKLDFTIILNETKF-WYQMIPLPH-FPTSKKYPLLDIVYAGPSCQ- 552
661 VNNRFGYKYPRLN---TLASLGYVVVV-IDNRGSGHGLFEGAFKYMQLIETDDOVE 716

```

```

Db 553 -----KDAIFELNMTATYLASTENIIVASFDRGSGYGQDKIMHAVNRRLGTFEVEDQIE 607
Qy 717 GLQYIASYDEITLDDRVGISHGMSYGYSLMAMQWSPIFRVALATGAVTLMIFDTQYV 776
Db 608 AARPD-SMGWGVDDRRIRAIWGMVSYGVTSMVLGAGSVFPGGLAVAFVSMEYDSVYT 666
Qy 777 ERYMGHP-DONEGQYGLGSVAM-OAEKPPSPNRLILHGLFENVHPAFHSILLSFLVR 834
Db 667 ERYMGLPRPQNLNDLYKNSYMSRAENF--KQVEYLILHGRADNVHVGQSAQISKALVD 724
Qy 835 AKRPDIQIYPOERHSIRVPESGEHYELHLILHYOE 870
Db 725 AGVDFOAMWYDEDDHGINSAPAHQIYTHMSHFQKQ 760

RESULT 2
ID DPp4_BOVIN STANDARD; PRT; 765 AA.
AC B81425; O8MMS8;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26) (Adenosine deaminase complexing protein) (ADCP-I)
DE (Activation molecule 3) (ACT3) (WC10).
GN DPp4 OR CD26.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphocytes;
RX MEDLINE=22067734; PubMed=12073152;
RA Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,
RT "Molecular characterization of bovine CD26 upregulated by a
RL staphylococcal superantigen."
RN Immunogenetics 54:216-220(2002).
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RC TISSUE=Thymus;
RX MEDLINE=22021197; PubMed=11981836;
RA Gliddon D.R., Howard C.J.;
RT "CD26 is expressed on a restricted subpopulation of dendritic cells in
RL vivo."
RN Eur. J. Immunol. 32:1472-1481(2002).
RN [3]
RP SEQUENCE OF 1-24.
RC TISSUE=T-cell;
RX MEDLINE=2148204; PubMed=11598101;
RA Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,
RT Naessens J., Bothach G.A.;
RL "Identity of activation molecule 3 on superantigen-stimulated bovine
RT cells is CD26."
RN Infect. Immun. 69:7190-7193(2001).
RN [4]
RP SEQUENCE OF 537-546.
RC TISSUE=Kidney;
RX MEDLINE=96293306; PubMed=9629661;
RA Ben-Shooshan I., Patroia A.H.;
RT "The CP-I subunit of adenosine deaminase complexing protein from calf
RL kidney is identical to human, mouse, and rat dipeptidyl peptidase
RT IV."
RN Comp. Biochem. Physiol. 119B:289-297(1998).
RP FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline. Binds and regulates the activity
CC of ADA.
CC -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC xcc, from a polypeptide, preferentially when xbb is Pro, provided
CC xcc is neither Pro nor hydroxyproline.

```

```

CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -1- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and
CC several immune system tissues.
CC -1- PFM: The soluble form (SDP) derives from the membrane form (MDP)
CC by proteolytic processing (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF461806; AAL67836.1; -.
CC EMBL; AY056834; AAL23628.1; -.
CC MEMPROS; S09_003; -.
CC InterPro: IPR002471; ProL_endopep_ser.
CC InterPro: IPR000379; Ser_estrs_site.
CC Pfam; PF00930; DDPV_N_term; 1.
CC PROSITE; PS00708; PEPTIDASE_S9; 1.
CC PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
CC Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
CC Transmembrane; Glycoprotein; Signal-anchor.
CC CHAIN 1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
CC (MDP).
CC CHAIN 38 765 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
CC (SDP) (By similarity).
CC FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 7 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC FT ACT_SITE 30 765 EXTRACELLULAR (POTENTIAL).
CC FT ACT_SITE 629 629 CHARGE RELAY SYSTEM (By similarity).
CC FT ACT_SITE 707 707 CHARGE RELAY SYSTEM (By similarity).
CC FT ACT_SITE 739 739 CHARGE RELAY SYSTEM (By similarity).
CC FT CARBOHYD 84 84 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 91 91 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 149 149 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 218 218 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 228 228 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 271 271 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 320 320 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 392 392 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 684 684 N-LINKED (GLCNAC...) (POTENTIAL).
CC SEQUENCE 765 AA; 88369 MW; E32165421F43E116 CRC64;

```

Query Match 12.0%; Score 566; DB 1; Length 765;
 Best Local Similarity 27.0%; Pred. No. 1.5e-33;
 Matches 191; Conservative 105; Mismatches 274; Indels 138; Gaps 29;

```

QY 206 PADPWI-----AFHSNDWISNVTREERLTGVNHELMAMERDASAGATVY 256
DB 148 PUNTWITSSVGHKLAIVWMDIYKXPNBDSQITWTG-----KDVITNGITDVI 201
QY 257 IQEE-FDRYSGYWCPKAEETPSGGKILRLIYEENDESEVEIHH---VTSPLMETRAD 312
DB 202 YEEEFVSAVSALMWSNSTF-----LAVAQFNDTEVPLIEFSFYSDLSQYKTKY 252
QY 313 FRYEPTGTANPKVTFKMSIMIDABGRITIDVTDKELIOPFELFEGVEYIARAGWTPECK 372
DB 253 IYPYAGAVNP--TIKFVWNISLSPINATSOQVPPGSVLI-GDHYLCVTVWTEER 309
QY 373 YAMSLIDRSQTRLQIVLISPELFPVBDVMERQGLISVDSVTPLLIYEBETDIDWYN 432
DB 310 IS-----LQMLRIQNY--STNDICDDYRSGTGRWIS 338
QY 433 IHDIVHVFQSHHEEL-----EFIFASECKTGFRHLKYITSLKSKYKRSRG 480

```

```

DB 339 SVGRQHI-----EISTGWGRFPAEPHPTSDGNS----FYKIIS--NEEGYKHI-- 383
QY 481 GLPAPSDFKCPKEIEAITSGEWEVLGRHGSNIQVDEVRLEYFEGT-KDSPLEHLLYVV 539
DB 384 -----CHFQDKRNCNCFITKGAMEVIG-----IEALTSDVLYIISNWKQMPARNLKYI 433
QY 540 SYNPEBEVYRLT-----DREYSHSCCISQCHDFEISKYSNQKNPCHVSLYKLSPEDDPT 594
DB 434 QLNDYKVTCTSELNPNDRCOQYVSFSQPAKYQRLRCSGRLP---LYTLHNSND-- 487
QY 595 CKTEKFWATIDAGPLPD-----YPPPELFSESTTGFLT-GMLYKPHDLQPKKYPT 648
DB 488 ---KEL--RVLENNSDLDVYLQVQMPKSLDPIHLLHGTFKWMILPPH-PDKSKRYPL 541
QY 649 VLFYGGPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 704
DB 542 LLEVYAGPCSQ-----KDAIFRLWATYLAATENIIVASFGDRSGYGQDKIMHAINR 595
QY 705 KMQQIEIDDOVEGLQYTLASRYPTIDLDVNGIHGMSYGGYSLMALMQRSDIFRYAAGAP 764
DB 596 RLGTFFVEDQIEATRFQF-SKMGFVDDKRIALMGMSYGYVTSVVLGAGSGVFFKGLIAYAP 654
QY 765 VTLMIFYDNGYTRRWNG--HPDNEGGYIIGSVAMQAEKRPSPRNLILLHGFLDENVHF 822
DB 655 VSKMEYDYSVYTRKNGLPPEPDNLSYRSTVMSRAENF--KQVEYLLTHGTADQVHF 712
QY 823 AHTSILLFLYRAGKPYDLQIYQERHSIRVPSGEGHYELHLHYOE 870
DB 713 QQAQISKALVDAGVDFQSMWVYTDDEHGIASSYRAHQHITYHNSHFLKQ 760

```

RESULT 3
 DDP4 HUMAN STANDARD; PRT; 766 AA.
 AC E27487;
 DT 01-APR-1992 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
 antigen CD26) (TPI03) (Adenosine deaminase complexing protein-2)
 DE (ADABP).
 GN DPP4 OR ADPC2 OR CD26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN NCBI_TaxID=9606;
 RC [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=92329551; PubMed=1352704;
 RA Mishumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of human dipeptidyl
 RT peptidase IV, a serine proteinase on the cell surface.";
 RL Biochim. Biophys. Acta 1131:333-336 (1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95012454; PubMed=7927537;
 RA Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
 RT "Genomic organization, exact localization, and tissue expression of
 RT the human CD26 (dipeptidyl peptidase IV) gene.";
 RL Immunogenetics 40:331-338 (1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=peripheral blood;
 RX MEDLINE=92325476; PubMed=1352530;
 RA Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,
 RA Dahlberg H.N., Schlossman S.F., Morimoto C.;
 RT "Cloning and functional expression of the T cell activation antigen
 RT CD26.";
 RL J. Immunol. 149:481-486 (1992).
 RN [4]
 RP ERRATUM.

FX MEDLINE=93j171637; PubMed=8094732;
RA Tanaka T.;
RL J. Immunol. 150:2090-2090(1993).
RN [5]
RN SEQUENCE FROM N.A.
RP TISSUE=Prostate;
RC MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,
RA Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley P.C., Hake S., Garcia A.M., Gay L.J., Hulys S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Witting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Sutterch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RN SEQUENCE OF 1-551 FROM N.A.
RP TISSUE=Colon;
RC MEDLINE=92i165847; PubMed=1347043;
RX Darnoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,
RA Tiroct P., Barbat A.;
RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like
RT colon cancer cell lines HT-29 and Caco-2. Cloning of the complete
RT human coding sequence and changes of dipeptidyl peptidase IV mRNA
RT levels during cell differentiation".
RL J. Biol. Chem. 267:4824-4833(1992).
RN [7]
RN SEQUENCE OF 545-766 FROM N.A.
RP TISSUE=Colon;
RC MEDLINE=91o24044; PubMed=1977364;
RX Darnoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;
RT "Isolation of a cDNA probe for the human intestinal
RT dipeptidylpeptidase IV and assignment of the gene locus DPP4 to
RT chromosome 2.";
RL Ann. Hum. Genet. 54:191-197(1990).
RN [8]
RN SEQUENCE OF 1-31 FROM N.A.
RP MEDLINE=96o67599; PubMed=7487939;
RX Boehm S.K., Gum J.R. Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific
RT regulation from a TATA-less GC-rich sequence characteristic of a
RT housekeeping gene promoter".
RL Biochem. J. 311:835-843(1995).
RN [9]
RN PARTIAL SEQUENCE.
RP TISSUE=Kidney;
RC MEDLINE=93j10468; PubMed=8096237;
RX Morrison M.E., Vijayasaradhi S., Engelstein D., Albino A.P.,
RA Houghton N.N.;
RT "A marker for neoplastic progression of human melanocytes is a cell
RT surface ectopeptidase.";
RL J. Exp. Med. 177:1135-1143(1993).
RN [10]
RN FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline. Plays a role in T cell activation
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer or heterodimer with seprase (FAP).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form.

```
CC CC -I- PWM: The soluble form (SPP) derives from the membrane form (MDP)  
CC by proteolytic processing.  
CC -I- SIMILARITY: Belongs to peptidase family S9B.  
CC -I- DATABASE: NAME=PROW; NOTE=CD guide CD26 entry;  
CC WWW:http://www.ncbi.nlm.nih.gov/prow/cd/cd26.htm".  
CC  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL, U13735; AAB60646.1; -.  
DR EMBL, U13710; AAB60646.1; JOINED.  
DR EMBL, U13711; AAB60646.1; JOINED.  
DR EMBL, U13712; AAB60646.1; JOINED.  
DR EMBL, U13713; AAB60646.1; JOINED.  
DR EMBL, U13714; AAB60646.1; JOINED.  
DR EMBL, U13715; AAB60646.1; JOINED.  
DR EMBL, U13716; AAB60646.1; JOINED.  
DR EMBL, U13717; AAB60646.1; JOINED.  
DR EMBL, U13718; AAB60646.1; JOINED.  
DR EMBL, U13719; AAB60646.1; JOINED.  
DR EMBL, U13720; AAB60646.1; JOINED.  
DR EMBL, U13721; AAB60646.1; JOINED.  
DR EMBL, U13722; AAB60646.1; JOINED.  
DR EMBL, U13723; AAB60646.1; JOINED.  
DR EMBL, U13724; AAB60646.1; JOINED.  
DR EMBL, U13725; AAB60646.1; JOINED.  
DR EMBL, U13726; AAB60646.1; JOINED.  
DR EMBL, U13727; AAB60646.1; JOINED.  
DR EMBL, U13728; AAB60646.1; JOINED.  
DR EMBL, U13729; AAB60646.1; JOINED.  
DR EMBL, U13730; AAB60646.1; JOINED.  
DR EMBL, U13731; AAB60646.1; JOINED.  
DR EMBL, U13732; AAB60646.1; JOINED.  
DR EMBL, U13733; AAB60646.1; JOINED.  
DR EMBL, U13734; AAB60646.1; JOINED.  
DR EMBL, M74777; AAA51943.1; -.  
DR EMBL, BC013329; AAH13329.1; -.  
DR EMBL, M80536; AAA52308.1; -.  
DR EMBL, X60708; CAA43118.1; -.  
DR EMBL, S79876; AAB35614.1; -.  
DR PIR, S24313; CDHU26.  
DR PDB, INIM; 07-JAN-03.  
DR MEROPS, S09_003; -.  
DR Genew; HGNC:3009; DPP4.  
DR MIM; 102720; -.  
DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; TAS.  
DR GO; GO:0005209; F:Plasma protein; TAS.  
DR GO; GO:0006955; P:Immune response; TAS.  
DR InterPro; IPR002469; DPPIV_N term.  
DR InterPro; IPR001375; Peptidase_S9.  
DR InterPro; IPR002471; Prol_endopep_ser.  
DR InterPro; IPR004379; Ser_estrs_site.  
DR Pfam; PF00930; DPPIV_N term; 1.  
DR Prosite; PS00708; PRO_ENDOPEP_SER; 1.  
DR Hydrolyse; Aminopeptidase; Dipeptidase; Serine protease;  
KM Transmembrane; Glycoprotein; Signal-anchor; 3D-structure.  
FT CHAIN 1 766  
FT FT  
FT CHAIN 39 766  
FT FT  
FT DOMAIN 1 6  
FT TRANSSEM 7 28  
FT FT  
FT DOMAIN 29 766  
FT ACT_SITE 630 630  
FT ACT_SITE 708 708  
FT ACT_SITE 740 740
```

FT	CARBOHYD	85	85	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	92	92	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	150	150	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	219	219	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	229	229	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	281	281	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	321	321	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	520	520	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	685	685	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	6	6	K -> R	(IN REF. 6).	
FT	CONFLICT	7	7	V -> I	(IN REF. 1).	
FT	CONFLICT	437	437	S -> I	(IN REF. 1).	
FT	CONFLICT	557	557	T -> I	(IN REF. 7).	
FT	CONFLICT	663	663	D -> E	(IN REF. 7).	

Query Match	11.2%	Score 525;	DB 1;	Length 766;
Best Local Similarity	26.3%	Pred. No. 1.5e-30;		
Matches 186;	Conservative 108;	Mismatches 277;	Indels 136;	Gaps 33

QY		206	PADDPMI-----AFHENDIMISIVIREERLTYVNHLEINMEDASAGAVFV	256
Dd		149	PNNQWMTWSVGKHLAYVANNDIYVKIEPMLPSRITWTG-----KEDIINGITDWM	202
QY		257	LQEB-PDRYSYGMCPKAETTPSGKILRIIYEENDESEVEIIH--VTSPLETRADS	312
Dd		203	YEEVEFSAYSALMSPNGTF-----LVAQCNDTEVPLETESFYSDSLQPKTVR	253
QY		313	FREYKTGANRKVTFFXNSEIMIIDAEGRLIDVIDKELIQFELLFGVEYIARAGTPECK	372
Dd		254	VPRYKAGVNPTVFKEV--VNTDSLSTYNATSIQITAPASMLI-GDHYLCDVTA----	306
QY		373	YAMSIDLSQTRQOVLISPFLIPVEDVMERQLIESVDSTYPLIIEETDIMN	432
Dd		307	-----TERISLOWL-----RIQNY--SYMDICDVSSGRM-N	338
QY		433	IHDIFHPVPOSHHEEIEFIFASECTGRRH-----LYKITSLIKESKYRSSGGLPAP	485
Dd		339	C----LVARGHIEMSTGTGWGRPRSPPHLTLDONSYKIIS-NEGGRHI-----	384
QY		486	SDEKCFIKEBIAITSGEWEVLGRHSNIQVDEVRLYFEQT-KDSPLEHLIYVSYPN	544
Dd		385	CYFOIDKKDCFTFKGTWEVIG-----IEALTSDPLYIISNEYKMPGGRNLTKL----	434
QY		545	GEVRLLDRGSHSCCIS-----QHCDPFISKYSOKRPH--C---VSLYKLSSPEDD	592
Dd		435	---QLSD--YTKYTCLSCSELNPRCOQYVSFSKEANYOLRCSGPGLPYTLTHSSVND	488
QY		593	PTCKTKEFMATILDSAGPLPDYTPBE--IPSFBSTGYTLGYMKYKHDLQPGKKYPTV	649
Dd		489	KGRVLED-NSALKRM--LQNVQMPRSKLDIILLNETGF-MYOMILRPH-FDKSKYVLL	543
QY		650	LFIYGSOVOLVNNRFKVGYKERLN--TLASLGYYVVV-IDNRSGCHGLKFEGAFYKX	705
Dd		544	LDVYAGCSQ-----KADTVFRLNWATYLASTEIIIASFDSGRSGVQGOKIMHAINRR	597
QY		706	MGOEIFDDOVBEGLOYLASRYDFIDLDRVGHGWSYGYSLMALMORSDFIRVALGAPV	765
Dd		598	LGFEEVEDJOIEAARQF-SKGQGVFNKRILAIMGWSYGYVTSMVYGSGBGVKCGIAPAV	656
QY		766	TLMFEYDTGTERFMG--HPDONOGYVLGSVANQAETFPBPBRLLIHGFLDENHFPA	823
Dd		657	SRMBEYDSVTERMGAPTREDNLDHKRNSTVMSRAEF--KOVEYLLIHGTADDNVHQ	714
QY		824	HTSILSFLVRAGKPLOLIPQBRHSIRPVDESAGEYELHLHLIOE	870
Dd		715	QSAQISALVDGVDPQMMYTDEDHGIASSTAQHITYHMSHTIKQ	761

DT 01-APR-1990 (Rel. 14, Last sequence update)
 DR 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) [T-cell activation
 antigen CD26] (gp110 glycoprotein) (Bile canaliculus domain-specific
 membrane glycoprotein).
 DE DPP4 OR CD26.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89123496; PubMed=2563382;
 RA Ogata S., Mitsumi Y., Ikehara Y.;
 RT "Primary structure of rat liver dipeptidyl peptidase IV deduced from
 RT its cDNA and identification of the NH2-terminal signal sequence as
 RT the membrane-anchoring domain.";
 RL J. Biol. Chem. 264:3596-3601(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88068516; PubMed=3479775;
 RA Hong W., Doyle D.;
 RT "cDNA cloning for a bile canaliculus domain-specific membrane
 RT glycoprotein of rat hepatocytes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).
 RN [3]
 RP SIGNAL-ANCHOR.
 RX MEDLINE=90338089; PubMed=1974258;
 RA Hong W., Doyle D.;
 RT "Molecular dissection of the NH2-terminal signal/anchor sequence of
 RT rat dipeptidyl peptidase IV.";
 RL J. Cell Biol. 111:323-328(1990).
 RN [4]
 RP SEQUENCE OF 281-302.
 RC TISSUE=Kidney;
 RX MEDLINE=94128239; PubMed=7905271;
 RA Iwaki-Bgawa S., Watanabe Y., Fujimoto Y.;
 RT "N-terminal amino acid sequence of the 60-kDa protein of rat kidney
 RT dipeptidyl peptidase IV.";
 RL Biol. Chem. Hoppe-Seyler 374:973-975(1993).
 CC -1 FUNCTION. REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
 CC POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
 CC PENULTIMATE RESIDUE IS PROLINE.
 CC -1 CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-[
 CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
 CC Xcc is neither Pro nor hydroxyproline.
 CC -1 SUBUNIT: Homodimer.
 CC -1 SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
 CC A SOLUBLE FORM.
 CC -1 PIM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
 CC BY PROTEOLYTIC PROCESSING.
 CC -1 SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC [a](#)
 CC EMBL; J04591; AAA41096.1; -;
 DR EMBL; J02997; AAA41272.1; -;
 DR PIR; A39914; A39914.
 DR MEROPS; S09.003; -;
 DR InterPro; IPR002469; DPPIV_N_term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR002471; Prol_endopep_ser.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF00930; DPPIV_N_term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;

Transmembrane: Glycoprotein, Signal-anchor.
 FT CHAIN 1 767 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
 FT CHAIN 37 767 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
 FT CHAIN 281 767 DIPEPTIDYL PEPTIDASE IV 60 KDA SOLUBLE FORM.
 FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 7 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 29 767 EXTRACELLULAR (POTENTIAL).
 FT ACT SITE 631 631 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 709 709 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 741 741 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 686 686 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 38 38 A -> R (IN REF. 2).
 FT CONFLICT 183 183 I -> T (IN REF. 2).
 FT CONFLICT 332 332 C -> V (IN REF. 2).
 FT CONFLICT 352 352 V -> D (IN REF. 2).
 FT CONFLICT 394 394 L -> F (IN REF. 2).
 FT CONFLICT 562 562 R -> Q (IN REF. 2).
 FT CONFLICT 624 624 R -> Q (IN REF. 2).
 FT SEQUENCE 767 AA: 88003 MW: AA17C0E6F0F4652 CRC64;
 Query Match 11.1%; Score 523.5; DB 1; Length 767;
 Best Local Similarity 26.0%; Pred. No. 28-30; Indels 165; Gaps 33;
 Matches 188; Conservative 104; Mismatches 266;
 206 PADPWI-----AFHSNDIWSNIVTREERLTIVHNEILANMEEDARSAGVATFV 256
 147 PNTQWIMSGEGHKLAVYWKNDIYVKI BPHLRIRIT-----STGENVTFNFINDMV 200
 257 LOEF-FDRISGYWCPKAEITTPSGCKIIRILEENDESEVEIIVHTSPMLETRADSRFY 315
 201 YEEELFGAYSLWM-----SPNGTFLLAYAFNDGVPLIEISFYSDSLQYPKTWIPIY 254
 316 PRTGANPKVTFFKMEIMIDAGRILIDVDEKLIQPFELFEGVEYIARAGTPEPKTAW 375
 255 PKAGVAVTAVKFTI--VTDSLSSTTTTTPMOITPAVAT-TGDHYLCDAVAVSDRIS- 310
 376 SILDRSQTRLQIVLISELFIPEVDYMERQRLIESYVDSTPLIYEETDIWINIHD 435
 311 -----LQWLRIRQNY--SYMAICDVKITLWV-NC-- 337
 436 IFHVPQSHERIEIFASEC-----KTGRRLKYITSIILKE 472
 338 -----PTQEHIEISATGWCGRFPAPBPHFTSDSSFYKIVSDQGYKH--ICQPKD 388
 473 SKYRSSGGLPAPSPDFKCPKIEEIA--ITSGEWEVLGHSNIOVDVRLRYVEPGT-KD 529
 389 RK-----PEQVTRFTTKAMEVI-----SIEMLTSDIYLTISNEYKE 425
 530 SPLHHLYIVSYNNGEVRTLRDGYSHSCIS-----QHCDFF--ISKYSNOKNPPC- 580
 426 MFGGNLYKI-----QLTD--HTMKKCHSCDLNBERCQYVSVLSKEAKYQQLQCR 474
 581 ---SYLYLSSPEDPTCKTEKEMATILDSAGPLPDYPPR---IFSESTTGFLYLM 634
 475 GPGLEPLYLHRSSTQKELRVED-NSALDKM--LQDVOMPSKQDLFIYLNTRF-WYOMI 530
 635 YKPHDLQPKKYPTVLFYGGPOVQVNNRFGVYKPYFLN--TLASGYVVV-INDRG 690
 531 LPH-FDKSKYPLIIDVYAGCSQ-----KADAPFLNMTATYIASFENIIVAFDRG 583
 691 SCHGLKREGAFKTKMGQIEIDQVEGL-QYIASYDITLDRVGHGMSYGGYISLMA 749

Db 584 SGYQGDKIMAHINKRLGLEVEDQIEAARQFL--RMGEVDSKRAVIMGMSYGGYTSWVL 641
 QY 750 MQRSDIFEVATAGAPVTLMEFYDGTGYRYNG--HPDQNEQGYVLSGYAMAQEKPPSPN 807
 Db 642 GSGSVFECGLAVAPVSMWEYDYDSYTERVWGLPPEPNDLHRYNSTMSRKNF--KQV 699
 QY 808 RLLIHGFLDENVHPARTSILSLVRAKCPDYDQIYQOERHSIRVPSGHEYLHLIHY 867
 Db 700 EYLLHGTADNVHFQQAQISKALVDAGVDFQAMWYDDEBHGIASTAHQHIYSHMSHF 759
 QY 868 LOE 870
 Db 760 LQQ 762
 RESULT 5
 ID_YEAS SCHPO STANDARD; PRT; 793 AA.
 AC 014073;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Putative dipeptidyl aminopeptidase C2E11.08 (EC 3.4.14.-).
 GN SPACUW4.08 OR SPAC2E11.08 (Fission yeast).
 OC Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetes; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN NCBI_TaxID=4896;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RC MEDLINE=21846401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou S., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Frazer A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voiciklaert G., Aert R., Robben J., Grymoprez B.,
 RA Welljens I., Vanstraelen E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Weiler H., Wambut R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Dada R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shipakowski G.V., Ussery D., Barrett B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
 CC vacuoles (By similarity)
 CC -I- SIMILARITY: Belongs to peptidase family S9B.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL031180; CAA20138.1; -
 DR PIR; T41703;

DR Genedb_Spombe: SPACUNK4.08; -
 DR InterPro: IPR002469; DDPiv_N term.
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR002471; ProI_endopep_ser.
 DR InterPro: IPR000379; Ser_estrs_site.
 DR Pfam: PF00930; DDPiv_N term: 1.
 DR Pfam: PF00326; Peptidase_S9: 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 DR Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
 KW Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
 FT DOMAIN 1 24 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 25 45 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT LUMENAL (POTENTIAL).
 FT ACT_SITE 46 793 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 647 722 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 722 722 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 755 755 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 101 101 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 246 246 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 303 303 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 324 324 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 336 336 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 377 377 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 384 384 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 407 407 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 535 535 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 761 761 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 793 AA; 91304 MW; 2087079797231463 CRC64;

Query Match 11.1%; Score 519.5; DB 1; Length 793;
 Best Local Similarity 26.3%; Pred. No. 4e-30;
 Matches 187; Conservative 113; Mismatches 256; Indels 155; Gaps 32;

212 IAFHSNDIMI-----SNIVTREERRLTY-----VHNELMNERDASAGVAFVLOEE-F 261
 183 LSFVYNNDIYVRKNGNV-----QRLTYGYDVFN-----GLTWIYEEVYL 225
 262 DRYSGYWCPCAKETTPSGGKILRIILYEENDESEVELIHV-----TSPMLERRAD 312
 226 SSPSTIWMSPDSD-----KIAPLKLNESEIPYHYPLVYALDELPLPEDVYNKMA 276
 313 FRYPKTGTANPKVTFKMSIMIDAEKRIIDVIDKELIQPELLFEGVYVIAAGWTPECK 372
 277 IYPPKGNPNPVSLEFVADLNSNASSNFMNHEPLAEP-----VVOQVLMVWNS----- 326
 373 YAMSILL-----DRSOTFLQIVLISPEL--FIPVEDVMER-----ORLISVPDSVTPLI 422
 327 ---SVLVQFTNNSTCITARLDELTKSIHYVKTCELEGWEVQASACMP--LNSILV 381
 423 YEETDIWIMIHIDFHVPOSHHEEIEFLPASECKTGFRHLYKITSILKSKYKSSGGL 482
 382 WENMSD--GYFDLALDDYNHLAIFP-----NGSS 410
 483 PAPSPKCIKEIEIATSGEMVLGHSNIOVDYRRLVYEGTQDSLEHLLVYVYV 542
 411 P-----IYVTSAMVYT--DGRPHIDGFGNVYPLATLKOSTERHLYVVS-L 454
 543 NGEVYTRLDRGYS---HSCCISOHCDPEISKYSQKNPCVSLYKLSPEDEPCKTKE 599
 455 DLEIYVIGITNDEDEGYSTSPSPEDFVLYNHGDVY---MGEIKTKQKDYCLSLIE 510
 600 FWATILDSAGPLPDYTPPEI---FSFESTGPTLYGMYLKRPHDLOPGKKYPTVLFTYGG 655
 511 TNSRLKQO---LSSITLPSVEYVKLTENDT--FNF--MERRRPNFEDVNNKAYVLFAYAG 564
 656 PVOVLVNNRFKGVKFFRLNTLASLG-----YVVVVINDRGSGCHGLKKEGAFYKMGQI 709
 565 PEGQOV-----AKIFRVDFOAYLASHDFEFTVTLDERGFGNGNARFYSVSSHLGEM 618
 710 EIDDOVEGIQYLASRYDFIDLRVGIHGSYGYLSIMALMORSDFIRVAVIAGAPVTLMI 769

Db 619 ESYDQQAQAKFMAD-LPFPDENHVGIGWMSYGYTLTKLT-ETQDVFSYGMANAPVTDNR 676
 QY 770 FYDTGYTERYMGHPDNEQGYLYGVAMQAEKFPSEPNLLHAGFLDENVHFAHTSILL 829
 Db 677 LVDSVYTERYMDLPQYNKSG-YKNSQIHDEYKF-KQLKXFFVAHGTGDGVHGHSHMLM 734
 QY 830 SFVLRAG-KPYDIQIYPERHSTRVPESEGEHFLHLYQENLGSRIAL 879
 Db 735 DGINLANCYNDMAVFPDSASHI---SYHNASTIYHRLSEWIGDALGRI 781
 RESULT 6
 DDP4_MOUSE
 ID DDP4_MOUSE STANDARD; PRT; 760 AA.
 AC P28843;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
 DE antigen CD26) (Thymocyte-activating molecule) (THAM).
 GN DPP4 OR CD26.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss; TISSUE=Thymus;
 RX MEDLINE=92129288; PubMed=1370813;
 RA Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
 RA Pierres M.;
 RT "cDNA cloning for mouse thymocyte-activating molecule. A
 RT multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
 RT subgroup of serine proteases.";
 RL J. Biol. Chem. 267:2200-2208 (1992).
 RN [2]
 RP REVISIONS.
 RA Marguet D.A.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B10.A; TISSUE=Liver;
 RX MEDLINE=95092780; PubMed=7999781;
 RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
 RT "Structure of the mouse dipeptidyl peptidase IV (CD26) gene";
 RL Biochemistry 33:15204-15214 (1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haich P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin A.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ushed T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mollahy S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Rahy U., Helton E., Kettelman M., Madan A., Young A.C., Shevchenko Y.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield J., Schin J.E., Jones S.J.M., Marra M.A., Small D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A., Small D.E.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [5]
 RP SEQUENCE OF 1-20.
 RX MEDLINE=91302787; PubMed=1712807;
 RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,

RA Bernard A.-M., Gorvel J.-P., Pierres M.;
 RT "evidence that thymocyte-activating molecule is mouse CD26
 (dipeptidyl peptidase IV).";
 RL J. Immunol. 147:447-454(1991).
 CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
 CC polypeptides having unsubstituted N-termini provided that the
 CC penultimate residue is proline.
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-
 CC xcc, from a polypeptide, preferentially when Xbb is Pro, provided
 CC xcc is neither Pro nor hydroxyproline.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
 CC a soluble form.
 CC -1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
 CC by proteolytic processing.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X58384; CAA41274.1; -.
 DR EMBL: U12620; AAA82213.1; -.
 DR EMBL: U12599; AAA82213.1; JOINED.
 DR EMBL: U12600; AAA82213.1; JOINED.
 DR EMBL: U12601; AAA82213.1; JOINED.
 DR EMBL: U12602; AAA82213.1; JOINED.
 DR EMBL: U12603; AAA82213.1; JOINED.
 DR EMBL: U12604; AAA82213.1; JOINED.
 DR EMBL: U12605; AAA82213.1; JOINED.
 DR EMBL: U12606; AAA82213.1; JOINED.
 DR EMBL: U12607; AAA82213.1; JOINED.
 DR EMBL: U12608; AAA82213.1; JOINED.
 DR EMBL: U12609; AAA82213.1; JOINED.
 DR EMBL: U12610; AAA82213.1; JOINED.
 DR EMBL: U12611; AAA82213.1; JOINED.
 DR EMBL: U12612; AAA82213.1; JOINED.
 DR EMBL: U12613; AAA82213.1; JOINED.
 DR EMBL: U12614; AAA82213.1; JOINED.
 DR EMBL: U12615; AAA82213.1; JOINED.
 DR EMBL: U12616; AAA82213.1; JOINED.
 DR EMBL: U12617; AAA82213.1; JOINED.
 DR EMBL: U12618; AAA82213.1; JOINED.
 DR EMBL: U12619; AAA82213.1; JOINED.
 DR EMBL: BC022183; AAH22183.1; -.
 DR MEROPS: S09.003; -.
 DR MGD: MGI:94919; DDP4.
 DR InterPro: IPR002469; DDPV_N_term.
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR002471; Prol_endopep_ser.
 DR InterPro: IPR000379; Ser_estrs_site.
 DR Pfam: PF00930; DDPV_N_term; 1.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 DR HydroLase: Aminopeptidase; Dipeptidase; Serine protease;
 KM Transmembrane; Glycoprotein; Signal-anchor.
 FT CHAIN 1 760
 FT FT
 FT CHAIN 37 760
 FT FT
 FT DOMAIN 1 6
 FT TRANSSEM 7 28
 FT FT
 FT DOMAIN 29 760
 FT ACT SITE 295 466
 FT ACT SITE 624 624
 FT ACT SITE 702 702
 FT ACT SITE 734 734
 FT CARBOHYD 83 83

FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 514 514 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 760 AA; 87436 MW; A5F644B6A3DF8 CRC64;
 Query Match 11.0%; Score 517.5; DB 1; Length 760;
 Best Local Similarity 25.0%; Pred. No. 5.3e-50;
 Matches 203; Conservative 107; Mismatches 306; Indels 197; Gaps 37;
 QY 131 MYSREBELRERRKIGTVGI-----ASYDYQGSGTFPLF-----OAGS 168
 DB 67 LYQENNNILLNAHNSSTFLENSTFESGCHSVSPDRFLVLENNYVQMKHSTASY 126
 QY 169 GIYHVXKDGPGFTQGPLRENLVETSCPNIRMDPKLCADPDMI-----AFHSND 219
 DB 127 NIYVNRK-----KOLITEKIPN-----NTQWITSPGKHLAYVWKND 165
 QY 220 IWSNIVTEERLTYVNHLEAMEDARSAGIATVLOE-PDRISGYWCKR----- 272
 DB 166 IYKVEPHLPSSRIT-----STGEENVLYNGITDWVEEVEFGAYSALWSPNNTFLAY 219
 QY 273 AETTPSGKILRLIYENDESEVEIITHVSPMLETRADSFRYPKTGTANPKYFKMSEI 332
 DB 220 AQFNDTGVPLEYSF-YSDS-----LQYKTVWIPPKGAVNPTKYFF--V 265
 QY 333 MIDABRIIDVIDKELIOPPEILFEGEVYIARAGWTPGKYA-----WSLLDRSQ 383
 DB 266 NIDSLSSSSAPLIQIPAPASVA-RGDHYLCVWMAETEIRISQMLRIRIQNVMAICDY 324
 QY 384 TRQIVLISPELPIPEDDVMERORLIESVPDVTLLIYEETDWINIHDFHVRPOS 443
 DB 325 DKTNLTWNCPS-----EQOHV-----EMSTGTGVG--RFRPAEPHF 358
 QY 444 HEEIEPIFASCKCTFRLHYKITSILKESKYRSSGGLPAPADF-KCPIKEIATITSGE 502
 DB 359 TSDGSSFYKISIDKQYKIHCHR-----PDKKDCITF-----ITKGA 395
 QY 503 WEYVLRHGSNIGQDEVRLVYFEGT-KDSPLEHLHYVSVYNDGEVTRLTRDGYSHSCCI 561
 DB 396 WEVI-----SIEALTSDDYLYISNOYKEMGCHNLKYI-----QLTD--HTNVKCL 439
 QY 562 S-----QHCDFFISKXSNQKNPH---C-----VSLYKLSPEDDPTCKTKEFMATIIDSAG 609
 DB 440 SCDLNPBRCCOYAVVSFSKEAKYQOLCQWGBPLTYLHRSSTDHKEKLVLE-----DNSA 493
 QY 610 ---PLDDYTPPE---IFSESTIGFTLYGMLYKPHDLQPKKYPYVLFYIGHQVQLVNN 663
 DB 494 LDBMLQDVQWPSKCLDFIVLNETRF-WYQMLLPFH-FDKSKKYPILLDVAAGCSQ---- 547
 QY 664 RFGVQKYPRLN---TLASLGIVVVV-IDNRGSGHRLGFKPGAKYKMGQJEIFDDVQGLQ 719
 DB 548 --KADASFRLNMAVTAISTENIIVASFDDGSGSYQDDKIMHANNRGLTELEVDDQIEAAR 605
 QY 720 YLASRDFIDLDVGVGHGMSYGYLSIMALMQSDIFRYAIAAPATLMIFYTGYTERY 779
 DB 606 QFP-KMGFVDSKVALWMSYGYTSMTVLGSSSGVFKGCIAYAPSRWYIYDSVTERY 664
 QY 780 MG--HPDQNEQGYGLGSVAMQAEKFPSEPNRLILLHGFIDENVHFAHSTILLSFLYRAG 837
 DB 665 MGLPIEDNLDHYRNSVTWMSRAHF--KQVEYLLIHGTADNVHFGQSAQISALVDAGV 722
 QY 838 PYLOLYIYQQRHSIRVPESGEHTEHLFLAYLOE 870
 DB 722 DFQAMWYTDDEHGIASTAHQHTIYSHMSHPLQ 755
 RESULT 7
 DAP2_YEAST

ID	DAP2 YEAST	STANDARD;	PRT;	818 AA.
AC	P19862;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DBAP B) (YSCV).			
GN	DAP2 OR YHR028C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
RN	NCBI_TaxID=4932;			
RN	[1]			
RX	SEQUENCE FROM N.A.			
RA	MEDLINE=89174571; PubMed=2647766;			
RA	Roberts C.J., Pohlig G., Rothman J.H., Stevens T.H.;			
RT	"Structure, biosynthesis, and localization of dipeptidyl			
RT	aminopeptidase B, an integral membrane glycoprotein of the yeast			
RL	vacuole";			
RL	J. Cell Biol. 108:1363-1373(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=S288C / AB972;			
RX	MEDLINE=94378003; PubMed=8091229;			
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,			
RA	Du Z., Favell A., Fulton L., Gatungu S., Geisel C., Kirsten J.,			
RA	Kucada T., Hillier E.J., Piter W., Thomson L., Langston Y.,			
RA	Luttrell P., Louis B.J., Macri C., Mardis E., Meneses S., Mouser L.,			
RA	Nham M., Rifkin L., Riley L., St Peter H., Trevasik E., Vaughan K.,			
RA	Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,			
RT	Vandin M.;			
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome			
RL	VIII.";			
RL	Science 265:2077-2082(1994).			
CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSSOME-LIKE			
CC	VACUOLE.			
CC	-I- SIMILARITY: Belongs to peptidase family S9B.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as their content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isd-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC				
DR	EMBL; X15484; CAA33512.1; .			
DR	EMBL; U10399; AAB68879.1; .			
DR	PIR; S46780; A30107.			
DR	MEROPS; S09.006; .			
DR	GSD; S0001070; DAP2.			
DR	GO; GO:0000329; C:vacuolar membrane (sensu Fungi); IDA.			
DR	InterPro; IPR002469; DPPIV_N term.			
DR	InterPro; IPR001375; Peptidase_S9			
DR	InterPro; IPR002471; prol_endopep_ser.			
DR	InterPro; IPR000379; Ser_estrs_site.			
DR	Pfam; PF00930; DPPIV_N_term; 1.			
DR	Pfam; PF00326; Peptidase_S9; 1.			
DR	PROSITE; PS00708; PRO_ENDOPEP_SER; 1.			
KW	Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;			
FT	Transmembrane; Glycoprotein; Signal anchor.			
FT	CYTOPLASMIC (POTENTIAL).			
FT	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)			
FT	(POTENTIAL).			
FT	DOMAIN	1	29	
FT	TRANSMEM	30	45	
FT				
FT	DOMAIN	46	818	
FT	ACT_SITE	679	679	
FT	ACT_SITE	756	756	
FT	ACT_SITE	789	789	
FT	CARBOHYD	63	63	
FT	CARBOHYD	79	79	
FT	CARBOHYD	110	110	
FT	CARBOHYD	139	139	
FT	CARBOHYD	372	372	
FT	CARBOHYD	392	392	

Query Match	Best Local Similarity	Matches 183, Conservative 95, Mismatches 250, Indels 177, Gaps 35;
FT CARBOHYD 421 421	N-LINKED (GLCNAC, . .) (POTENTIAL) .	
FT CARBOHYD 738 738	N-LINKED (GLCNAC, . .) (POTENTIAL) .	
FT CONFLICT 83 83	Q -> H (IN REF. 1) .	
FT CONFLICT 125 125	S -> N (IN REF. 1) .	
FT CONFLICT 182 188	FEELONE -> LRLEET (IN REF. 1) .	
FT CONFLICT 200 200	D -> N (IN REF. 1) .	
FT CONFLICT 366 375	TANVARDSS -> DFRGKERKP (IN REF. 1) .	
FT CONFLICT 808 818	AKRPAQDFQVK -> QSVLSGNGLNELLTIYSSSHROIHKT PSYHTMYI (IN REF. 1) .	
SEQUENCE 818 AA; 93404 MM; 318P450445375BD3 CRC64;		
Query Match	10.3%; Score 484.5; DB 1; Length 818;	
Best Local Similarity	26.0%; Pred. No. 1.66-27;	
Matches 183, Conservative 95, Mismatches 250, Indels 177, Gaps 35;		
QY 212 IAFHSNDIWMISNVTREERLTYVHN-----IANMEDARASGAAVFLQOE-FDRYSG 266		
DB 201 IAYVODNNIYSAISKKTIRA-VITNDGSGFLNGKPD-----WVYEEEFEDDKA 250		
QY 267 YVMPCKAETTPSGKILRIILEENDESEV-EIHVTSPLMETRAD-----SPRYKGT 319		
DB 251 AMSEPTGD-----YLAFLKIDSEVGEFT--IPYVQEKDIYPEMRSIKYPSG 298		
QY 320 TANPKVTFKSEIMIDAEGRIIDVDELQIP-FELLFGEVYIADAGWPEEKVAMS 377		
DB 299 TPNP-----HAELWYWSKMGDTSPHPRISGNKKQGSLLITEVTWVNG-----NV 343		
QY 378 LL-----DSQIRQLVILSPELFIVEDDVMERQRLISVDSVTLPIIYETTDIWINI- 433		
DB 344 LVKTTDSSDLITLFFLL-----DLIAKTSNVVNN-----ESSNGMWEIT 383		
QY 434 HDIFHV-----PQSHHEIEIFIPASCKTGFRLLYKITSILKSKYKSSGGLPARSDF 488		
DB 384 HNTLFIPIANETFDPRPHGYVDILPIG---GYNHL-----AFENSN---SHY 425		
QY 489 KCPIKEIAITSGMEVLGRHGSNIQVDEVRRLVYFEGTQSPLEHLLYVSVYNPGEVT 548		
DB 426 K-----TLTEGKMEV---NGPLARDSMENHLFISTRKSSTERHVIYVILDRSPREIT 475		
QY 549 RLTRRGYS--HSCCISQHCDFPISKYSNQNPH-----C-----VSLYKL 586		
DB 476 EVIDTISBDGYVDVSFSSGRFRGLITYGPPVYQKIVDFHRSRAEKCDKGNVLGSKSYLH 535		
QY 587 SSPEDDPTCKTKERWATILDSAGLPIPYTPPELFSPESTGFLTYG--MLYKHHDLP-- 642		
DB 536 EKNE-----VLKTI--LEBYAVPRKSPFELNLGKDEFGDILVNSYEILPND 580		
QY 643 -----GKYPYTVLFIYGGPOVOLVNNRKFVKYKFFLN-----TLASIGVYVVDNRSGH 693		
DB 581 FDETLSDHYVFFPRAYGSPNSQY-----VKTFSGFNEVVASQMLAIYVVDGKGTGF 634		
QY 694 RGLKFESGAFYKMGQCIIFIDQVEGLQYLASRYDFIDLDRVIGHSYGGLTLMALMORS 753		
DB 635 KGQDFSLVLRDRLGDVEARDQISAAISLYGS-LTFVDPQKISLFGWSYGGLTLKTEKDG 693		
QY 754 D-IFRVALIAPATLWTFYDTGTERMGHPDQNEGYIIGSV-----AMQAEFPSEPNR 808		
DB 694 GRHKYIGMSAEPYVDMKFDYSVTEREMHPDENFGYESSYHNTAL-----AQANR 747		
QY 809 LLLHLGFLDENVFAHTSILSLFLVAG--KPYLQIYQPGRHSIR 852		
DB 748 FILMHGCGDNVHFQNSIKFLDLLDNGVENVYVHVPFSDHSIR 792		
RESULT 8		
SEPR MOUSE		
ID SEPR MOUSE STANDARD; PRT; 761 AA.		
AC P97321;		
DT 28-FEB-2003 (Rel. 41, Created)		
DT 28-FEB-2003 (Rel. 41, Last sequence update)		
DT 15-SEP-2003 (Rel. 42, Last annotation update)		
DE Sepriase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral membrane serine protease).		


```

Db 552 -----VAAMWITTLAKKEGIVIALVDORGAFOGDKFLHAYVKRLGYVEEDQLTAVRK 606
QY 721 LASEYDFIDLDRAVGHGMSYGYSLMALMQRSDIFRAVAGAPVTLIMFYDTYTERYM 780
Db 607 FL-EMGFIDEERIAMWMSYGYVSSIALAGTGLFKCIAGVAVSSMEVYASISERFM 665
QY 781 GHP--DQNEQGYIGSVAMQAEKPPSEPNRLLLHGLFDENAFHAFSTILISFLVRACKP 838
Db 666 GLPKDNLEHYKNKSTWARAERYFRND--YLLIHGTADNVHFQNSQAIKALVNAQVD 723
QY 839 YDLQIYPOERHSIRVPESGEYELHLHLYOE 870
Db 724 FQAMWYSDQNHGISGSSQNHLYTHMTFLKQ 755

RESULT 9
ST13_YEAST
ID ST13_YEAST STANDARD; PRT; 931 AA.
AC P33894;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase A (EC 3.4.14.-) (DPAp A) (YSCIV).
GN STE13 OR YCII OR YOR219C OR YOR50-9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9506382; Pubmed=7975897;
RA Anna-Artiola S.S., Herskowitz I.;
RT "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl
RT aminopeptidase."
RL Yeast 10:801-810(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=AB320;
RA Flanagan C.A., Thorner J.;
RT "STE13."
RL (in) Getting M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.);
RL Guidebook to the Yeast secretory pathway, pp.1-1, Oxford University
RL Press, Oxford (1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=5288C / FY1679;
RX MEDLINE=9643797; Pubmed=8840505;
RA Galisson F., Dujon B.;
RT "Sequence and analysis of a 33 kb fragment from the right arm of
RT chromosome XV of the yeast Saccharomyces cerevisiae."
RL Yeast 12:877-885(1996).
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC ALPHA-FACTOR PRECURSOR.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC VACUOLES.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L21944; AAA35119.1; -
CC EMBL; U08230; AAA17897.1; -
CC EMBL; X92441; CA53182.1; -
CC EMBL; Z75127; CA99437.1; -
CC FIC; A49737; A49737.
CC MEROPS; S09.005; -
CC SGD; S0005745; STE13.

```

```

DR GO; GO:0005802; C:Golgi trans-face; IDA.
DR GO; GO:0004177; F:aminopeptidase activity; IDA.
DR GO; GO:0007323; P:peptide pheromone maturation; IDA.
DR InterPro; IPR002469; DPIP N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR Pfam; PF00930; DPIP N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; FALSE NEG.
KM Hydrolyase; Aminopeptidase; Dipeptidase; Serine protease;
KM Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
FT DOMAIN 1 119
FT TRANSMEM 120 140
FT DOMAIN 141 931
FT ACT_SITE 785 785
FT ACT_SITE 863 863
FT ACT_SITE 896 896
FT CARBOHYD 377 377
FT CARBOHYD 814 814
SQ SEQUENCE 931 AA; 107200 MW; 81AF70094093C023 CRC64;

Query Match 9.9%; Score 465; DB 1; Length 931;
Best Local Similarity 21.5%; Pred. No. 5.1e-26;
Matches 218; Conservative 150; Mismatches 360; Indels 268; Gaps 41;

QY 6 ETEQLGVEIFETADCEENIESQDRPKLEPPVERYSQKLADTRKHYGMMAKAPH 65
DB 43 QSNENGHNTIMEIRFETETITVTDVPO-TPLQEQYS-----MRPRE 83
QY 66 DFMFVKANDDPGPHSDRIYLAWS-----GENRENTLFYSEIPIKTRINRAVILMS 115
DB 84 SPQE---NDIENQHHTSPFSVAKFNRMGBMSLPEKRSVIVVETLLALSVLVLLIIP 140
QY 116 WKPL-----LDLFOATL-----DYGM 131
DB 141 SKLLPTKTRPKTSAGDSSLGKRSFSLIENVINGDFALPEDTFHFIDPQRLGQDSBGL 200
QY 132 YSRBEELIRERKRIG-TVGIASVDYHQSGTFLQ----- 165
DB 201 YFTTKEIDGHTNFIKALQPEFTEVNLGKRFYEGVEFVSIVQVINYKLDKLIETNLE 260
QY 166 ----AGSGIYHVKDGPGQFTQOPLRPNLVETSCPNIRMDPK-----LCPAD 208
DB 261 SEFRHSSKGFYWKD-----INTGIEPILPEKSDDNVELGLSKLSYAHSPAP- 309
QY 209 EDWIAFIHSDIIVISNIVTREERLTY-VNELANNEEDARKSAGVATFYQEE-FDRYSG 266
DB 310 YNYIYFVEYENLFLQOVNSGVAKVTEDGSKDIFNAKPD-----WIYEEVLAQDA 361
QY 267 YWMCPRAEETPSGKILRLIYENDESEVRIIVTSPMLSTRADS-FRYPKGTAPKV 325
DB 362 IWW-----APDSKAVPARFNDTSVDDIRLKNYTN-MNEAVISDTIKYKPEGFQNP 414
QY 326 TFMASIMIDAEGRIIDVIDKELIQPEFILLFEGEVYIARAGWTPBEKTYASILDRSQTR 385
DB 415 DL---FLVNLQNGIIYSINTG-GQKDSILYNG-----KWISDPFEREL-TDRSKI 461
QY 386 IQVVLISPELFTVEEDWNERQQLISVDPSTPLIYETTDIWI-NIHDIHFVQSH 444
DB 462 LDVKYV-----DISSQML-TVNTMSNLF-----NGWTEKTDIISIPKPE 503
QY 445 EEEIE--FTFASCKTGFPHLYKITSILKESKYKRSSGGLPARSDPKPIKEEIALTSGE 502
DB 504 IAKMDGYIDIHADSGSFHLFYFVF-----AKEPILQLKGN 542
QY 503 WEYLGRH--GSNIQVDEVRRLVYFEGTKDPSLEHHLVYVSYV-----NPEGV 548
DB 543 WEYTGNGIYGEYETD---TIFFANEIGWMSQHLYSILSTSTQNTQNTQSLQNP----- 594
QY 549 RLIDRGYSHSCISQHCDFPIKYSNQKP-----HCYSLYLSSBEDD 552
DB 595 --SDKIDFYDFELSSSARAISKRLGPTPIYVAGPLTRVLNVAELIHDSITLQLTDE-- 650

```

QY 593 PCKTKKEFWATILDSAGLPDPYTPPELFSFESTT---GFTLYGMLYKPHDLQPGKPYTV 649
 Db 651 ---KFKK-----KIKYDLP-ITSYKTMVLDDVDEINYEIRKIPANINPKKKYDIL 696
 QY 650 LFIYGGPVOLVNNRFKGVKFRNLNTASLGIVVIVINRSGCHGLKFECAFXYMGI 709
 Db 697 VNIVGGPSSQFTT--KSSLAFEQAVVSGLDVIYQIPEPRGTGGGWSFRMARXKGYW 754
 QY 710 EIDDOVE-GLQYLSARYDFIDLDYRGIGMSYGYLSMAL-MORSDFRVAIGAPVTL 767
 Db 755 EPRITETVTKXFKFIQNSQIHIDESKIALMGWSYGFSTLKTYELNDGTFKXAMAVAPVTN 814
 QY 768 WIFPDGTERYMGHPDQNEQGYIGYVAMQAEKFPSPNNRLLHGLFDENVHFAHTSI 827
 Db 815 WTLVDSYTERYMQPSENHEGYFEVSTIQNFKSPES-LKRFLIVHGTEDDNVHLQNTFR 873
 QY 828 LLSFLVRAG-KFYDQIYPERHSTRVPESEGHYTHLHLHLOENLGRIALKVI 882
 Db 874 LVDQLNLGLTNYDMHIFPDSDSIRYHNAQRIYFQKLYWLRDAFAERFDNTEVL 929

RESULT 10
 SEPR_HUMAN STANDARD, PRT; 760 AA.
 ID Q12884; Q00139; Q99998; Q9UID4;
 AC Q12884; Q00139; Q99998; Q9UID4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral membrane serine protease) (170-kDa melanoma membrane-bound gelatinase).
 GN PAP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_TaxId:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Fibroblast;
 RX MEDLINE=94261645; PubMed=7911242;
 RA Scanlan M.J., Raj B.K.M., Calvo B., Garin-Chesa P., Sanz-Moncasi M.F., Healey J.H., Old L.J., Rettig W.J.;
 RT "Molecular cloning of fibroblast activation protein alpha, a member of the serine protease family selectively expressed in stromal fibroblasts of epithelial cancers.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Melanoma;
 RX MEDLINE=97388251; PubMed=9247085;
 RA Goldstein L.A., Gherzi G., Pineiro-Sanchez M.L., Salomone M., Yeh Y., Flessate D., Chen W.-T.;
 RT "Molecular cloning of seprase: a serine integral membrane protease from human melanoma.";
 RL Biochim. Biophys. Acta 1361:11-19(1997).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND 511-518.
 RC TISSUE=Melanoma;
 RX MEDLINE=97218181; PubMed=9065413;
 RA Pineiro-Sanchez M.L., Goldstein L.A., Dotti J., Howard L., Yeh Y., Chen W.-T.;
 RT "Identification of the 170-kDa melanoma membrane-bound gelatinase (seprase) as a serine integral membrane protease.";
 RL J. Biol. Chem. 272:7595-7601(1997).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Melanoma;
 RX MEDLINE=20112818; PubMed=10644713;
 RA Goldstein L.A., Chen W.-T.;
 RT "Identification of an alternatively spliced seprase mRNA that encodes a novel intracellular isoform.";

RL J. Biol. Chem. 275:2554-2559(2000).
 RN [5]
 RP SEQUENCE OF 192-208; 220-240 AND 510-521.
 RX MEDLINE=94327249; PubMed=7519584;
 RA Rettig W.J., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M., Garin-Chesa P., Healey J.H., Old L.J.;
 RT "Fibroblast activation protein: purification, epitope mapping and induction by growth factors.";
 RL Int. J. Cancer 58:385-392(1994).
 CC -1- FUNCTION: May have a role in tissue remodeling during development and wound healing, and may contribute to invasiveness in malignant cancers.
 CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and type IV collagen, but not native type I or type IV collagen. Does not cleave laminin, fibronectin, fibrin or casein.
 CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is inactive.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell surface lamellipodia, invadopodia and on shed vesicles.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=L;
 CC IsoId=Q12884-1; Sequence=Displayed;
 CC Note=Major isoform;
 CC Name=2; Synonyms=S, Truncated;
 CC IsoId=Q12884-2; Sequence=VSP_005367;
 CC -1- TISSUE SPECIFICITY: Fibroblast-specific.
 CC -1- INDUCTION: In fibroblasts at times and sites of tissue remodeling during development, tissue repair, and carcinogenesis.
 CC -1- PTM: N-glycosylated.
 CC -1- PTM: The N-terminus may be blocked.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U09378; AAB49652.1; -;
 DR EMBL; U76833; AAC51668.1; -;
 DR EMBL; AF007822; AAF21600.1; -;
 DR MEROPS; S09.007; -;
 DR GeneW; HGNC:3590; PAP.
 DR MIM; 600403; -;
 DR InterPro; IPR002469; DPP4V N term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR002471; Prol_endopep_ser.
 DR InterPro; IPR000379; Ser_estrs site.
 DR Pfam; PF00930; DPP4V N term; 1.
 DR Pfam; PF00325; Peptidase_S9; 1.
 DR PROSITE; PS00708; PRO_ENDOPEP_SRP; 1.
 KM Hydrolase; Protease; Serine protease; 1. Transmembrane; Signal-anchor;
 KM Glycoprotein; Alternative splicing.
 FT DOMAIN 1 4
 FT TRANSMEM 1 25
 FT FT 26 760
 FT ACT_SITE 624 624
 FT ACT_SITE 702 702
 FT ACT_SITE 734 734
 FT CARBOHYD 49 49
 FT CARBOHYD 92 92
 FT CARBOHYD 99 99
 FT CARBOHYD 314 314
 FT CARBOHYD 679 679
 FT VARSPLIC 1 521
 FT FT 207 207
 FT FT 229 229
 FT CONFLICT 354 354
 FT P -> A (IN REF. 2).
 FT K -> T (IN REF. 2).
 FT R -> T (IN REF. 2).
 FT CONFLICT 207 207
 FT CONFLICT 229 229
 FT CONFLICT 354 354


```

Db      367 EMTPTDDPRMREY-VITWVKATSTKVAWN---LSRQN-----VS----- 404
Qy      405 ERQRLIESVDSVPLII--YEETDIWINIHDIHFVFPQSHHEEIEFIFASECKTGFRH 462
Db      405 -----ILTLCDATVCTKKHDESEAML-----HRQNEPFFVS---KDG-RK 443
Qy      463 LYKITSILK--ESKYKRSGLPAPSDFKCPIKEEIAITSGEWEVLGRHSGNIQVDEVAR 520
Db      444 FFFVVAIPQGGGKRYHITVSSSQPNSSNDNIQ---SITSGMDVT---KILSDERKS 496
Qy      521 LVYFEGTDSPLFHHLYVSVYNPEVTR-----LITRGVSHSCCISCHCDFIFISKYSN 574
Db      497 QIYFSTEDLPERRQLYASATV--GSFNKQCLSDLVNCTVFSASFSGADFFLLKCEG 554
Qy      575 QKNPFCVSLYKLSSEDDPTCKTE-FWATILDSAGPLPDYPPPIFSFESTGTGLXGM 633
Db      555 PGVP-TVSVHNTTDDKKMFDELTNEHVQKSIDRMKPEYKITE-----TDDYNLIQ 607
Qy      634 LYKPHDLPQKKYFPVLFYGGPVQVLYNNRPKGVYFRLLTASLGYVVVVINDRSGCH 693
Db      608 ILKPAFTFDTAHYPLLLVVDGTPGQSQAEEK--AVTWETWVSSHGAVVVKCDGRSGF 665
Qy      694 RGLKEGAFKYMGOEIIDDOVEGLQYLASRYDFLDRVGHGMSYGGYLSIMLMQRS 753
Db      666 QSTRLLHEVRRRLGSLLEKQMEAVKVM--KEPYDKTRVAVFGDYGYLSTYLLPAKG 724
Qy      754 D-----IFRVALAGAVTLTMIPTGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPRL 809
Db      725 DQGAVFSGSALSTIDTFKLYASAFSERYLGHLDNRRAHYMAVARVSL--EGQGF 782
Qy      810 LLLHGFLENNVFAHTSILLSFVPAKPYDQIYPOERHSIRVSEGEHYELHLLHYLQ 869
Db      783 LVHATDADEKIHFORLAELITQLIKGKANYSLQIYPPDESHVSSAALQHLHRSITLGFV 842
Qy      870 E 870
Db      843 E 843

RESULT 12
DPP6_MOUSE STANDARD; PRT; 804 AA.
ID DPP6_MOUSE OSQW2; 092219;
AC 092218; OSQW2; 092219;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE (Dipeptidyl aminopeptidase-like protein) (DPPX).
GN DPP6 OR DPP-6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RC MEDLINE=99030650; PubMed=9811881;
RA Hough R.B., Lengeling A., Bedian V., Lo C., Bucan M.;
RT "Rump white inversion in the mouse disrupts dipeptidyl aminopeptidase-
RT like protein 6 and causes dysregulation of kit expression.";
RT Proc. Natl. Acad. Sci. U.S.A. 95:13800-13805(1998).
CC -!- FUNCTION: May be involved in the physiological processes of brain
CC activity. Has no dipeptidyl aminopeptidase activity. The lack of
CC for the serine residue in the proposed catalytic triad (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- SIMILARITY: Belongs to peptidase family S99.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).

```

CC      EMBL; AF092507; AAC97366.1; -
CC      DR      EMBL; AF092506; AAC97365.1; -
CC      DR      EMBL; AF092505; AAC98381.1; -
CC      DR      MEROPS; S09.973; -.
CC      MDI; MGI:94921; DPP6.
CC      DR      InterPro; IPR002469; Dipep. N term.
CC      DR      InterPro; IPR001375; Peptidase S9.
CC      DR      InterPro; IPR000379; Ser. setre site.
CC      DR      Pfam; PF00930; DPPV N term; 1.
CC      DR      Pfam; PF00326; Peptidase S9; 1.
CC      TRASNMBM; Glycoprotein; Signal-anchor.
CC      FT      TRANSMBM 1 34 CYTOPLASMIC (POTENTIAL).
CC      FT      TRANSMBM 35 55 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC      FT      TRANSMBM 56 804 EXTRACELLULAR (POTENTIAL).
CC      FT      TRANSMBM 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      TRANSMBM 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      TRANSMBM 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      TRANSMBM 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      TRANSMBM 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      TRANSMBM 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      TRANSMBM 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      TRANSMBM 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      TRANSMBM 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      TRANSMBM 440 455 MISSING (IN REF. 1; AAC98381).
CC      FT      TRANSMBM 638 638 O -> P (IN REF. 1; AAC97365).
CC      SQ      SEQUENCE 804 AA; 91260 MW; 09CFCE7AD8A7168 CRC64;

Query Match 9.3%; Score 435; DB 1; Length 804;
Similarity 23.0%; Pred. No. 6.5e-24;
Matches 178; Conservative 127; Mismatches 322; Indels 148; Gaps 30;

Qy      138 LLRERKRGITGVIGASYDHQSGTFLF-----QAGSGIYHYVDGSGPGTQQLRPN 189
Db      116 VLIQKKIESRLAIRYELSPDXEYVLFVSYNVEPYQSHHTGYVLSKIPHGDPQSLDPE 175
Qy      190 LVETSCPNIRMDPLCPADPDWIAFIHSNDIWNISNITREERRLLTYHNELANNEEDARS 249
Db      176 VSNAKLQYAGWGPK-----GQQLFIFENNIIYCAHVQKQAIRV-----VSTGKEGVIY 224
Qy      250 AGVATFVLOER-FRYSGYWMCPEKATTPSGCKLRLIYEENDESEVEIIV---TSMVL 305
Db      225 NGLSDWLYEBEILKSHIAHMSDPG-----TRLAVATINDSRVPLMELPYTGSVY 275
Qy      306 ETRRADSFRRYKGTGTANPKYTFKMSIMIDAGRIIDVIDKELIQPEELFEGVEYIARA 365
Db      276 PT--VKPYHYKAGSENPISLH-----VIGLNGPTHLD--EMMPDDPRRREY-YITMV 325
Qy      366 GWTEBGKYA--WILLDRSQRLQIVLSPPLFIVEDDVNERQLIESVDSVTPPLII- 422
Db      326 KWATSTKVAATVW--LNRQW-----VS-----LITLCDATVGVCTK 359
Qy      423 -YEETDIWINIHDIHFVFPQSHHEE-----EFIFASECKTGFR-HLYKITSILKSKX 474
Db      360 KHEBESAWLH-----KQNEPFFVSQDGRKFFVRAIPQGGKRYHIT--VSSSQ 408
Qy      475 YKRSGLPAPSDFKCPIKEEIAITSGEWEVLGRHSGNIQVDEVARLVYFEGTDSPLFH 534
Db      408 PNSSNDNIQ-----SITSGMDVT---KILSDERKNKIYFSTEDLPERR 451
Qy      535 HLYVSVYNPEVTRLLDRGVSHSCCISCHCDFIFISKYSNOKNPHCVSLYKLSSEDDPT 594
Db      452 HLVSANTVDDPNRQL-----SCDLVENCYVASFSHMD--FFLLCEGP-GVPT 500
Qy      595 C-----KTEFWATILDSAPLPDYTPPEPIFSFESTGTGLXGLYKXPH 639
Db      501 VYHNTTDDKKRMPLDLENEVQKAIANDRMKPEYKITEV-----EDYSLPMQILKPA 554

```


RESULT 14
 ID DP6 HUMAN STANDARD: PRT; 865 AA.
 AC P42558;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidyl aminopeptidase 6) (Dipeptidyl peptidase IV like protein)
 DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
 GN DPPX.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
 RC TISSUE=Hippocampus;
 RX MEDLINE=93372805; PubMed=8103397;
 RA Yokotani N., Doi K., Wenthold R.J., Wada K.;
 RT "Non-conservation of a catalytic residue in a dipeptidyl
 aminopeptidase IV-related protein encoded by a gene on human
 chromosome 7.";
 RL Hum. Mol. Genet. 2:1037-1039(1993).
 CC - FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 CC ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
 CC - SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC - ALTERNATIVE PRODUCTS: Type II membrane protein (Probable).
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC IsoId=P42658-1; Sequence=Displayed;
 CC Name=DPPX-S;
 CC IsoId=P42658-2; Sequence=VSP_005365;
 CC - TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
 CC - SIMILARITY: Belongs to peptidase family 59B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M96859; AAA35760.1; -;
 DR EMBL; M96860; AAA35761.1; -;
 DR PIR; I54331; I54331.
 DR MEROPS; S09.973; -;
 DR GeneW; HGNC:3010; DPP6.
 DR MIM; 126141; -;
 DR GO; GO:0008239; F:dipeptidyl-peptidase activity; TAS.
 DR InterPro; IPR002469; DPPV N term.
 DR InterPro; IPR003375; Peptidase S9.
 DR InterPro; IPR003379; Set_estrs_site.
 DR Pfam; PF00930; DPPV N term; 1.
 DR Pfam; PF00326; Peptidase S9; 1.
 DR Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 DR Transmembrane; Glycoprotein; Cytosolic (POTENTIAL).
 DR Transmembrane; Glycoprotein; Cytoplasmic (POTENTIAL).
 DR Transmembrane; Glycoprotein; Signal-anchor (TYPE-II MEMBRANE PROTEIN)
 DR Transmembrane; Glycoprotein; Signal-anchor (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 117 865
 FT CARBOHYD 173 865
 FT CARBOHYD 319 865
 FT CARBOHYD 404 865
 FT CARBOHYD 471 865
 FT CARBOHYD 535 865
 FT CARBOHYD 566 865
 FT CARBOHYD 813 865
 FT VASPLIC 1 81

FT KLGPRAQAAPRRERGAGGAGRFRFOYGRSDDEED
 FT -> MTTAKEPSAGSKSVQOQEOE (in isoform
 FT DPPX-S)
 FT /FTID=VSP_005365.
 SQ SEQUENCE 865 AA; 9758 MW; 14B1AE0E024464B CRC64;
 Query Match 9.1%; Score 427; DB 1; Length 865;
 Best Local Similarity 23.4%; Pred. No. 2,8e-23;
 Matches 172; Conservative 121; Mismatches 299; Indels 144; Gaps 30;
 171 YHVDGGGQGGTQGPRLNLTSCPNIRMPKCPADPMIAFHSDIWSNIVTSEE 230
 218 YVVLSTKIPHGDPQSLDPEEVSNAKLQYAGMGPK-----GQQLIFIFENNIIYCAHVQKA 272
 231 RRLTYVHNEILNMEEDARASAGVATFVLOE-FDRYSYGMPCPKAETTSGGKIIRILYEE 289
 273 IRV-----VSTGKGVITVNGLSJWLVEELIKTHIAIWMSPDG-----TLAAVA 317
 290 NDSEVEELIHY-----TSPMLETRRADSFRRYPKTGTANPKYFKMSEIMIDAEGLIDVIDK 346
 318 INDSRVPIMELPTVTSIYPT--VKPHYPRKASGNPISLH---VIGLNPHTDL--- 368
 347 ELIOPFELIFGVEYIARAGTTPGKYA--WSILDSQTRLQVLIPELFIYVEDDVM 404
 369 EMPEPDDPRKREY-YITWVKWATSTKVAVTW---LNRAQN-----VS----- 406
 405 ERQLIESVPSTVPLIIT--YEETDIWINIHDFHVPQSHFEI-----EFIFASEC 456
 407 -----ILTLCATITGVCTCKKHDESEANLH-----RQNEHYVPSKDGKFFITAI 452
 457 KTFGR-HLYKITSILKSKYRSSGGLPAPSPKCPKIEIATISGEMVILGRHGSNIQV 515
 453 PGGRGKGFYHIT--VSSSQPMSNDIQ-----SISGMDVY---KILAY 493
 516 DEVRLIYFEETKPSLEHLYVYSYVNPGEVTLTRDGRYSH---SCCISQHCDFPISKY 572
 494 DEKNKIKYFISTEDLPERRQLYSAN-----TEGNFRQCLSCDLVENCITYFSASF 543
 573 SNQKPHCVSLYKSSP-----EDDPCTKEP-----WATILDSAGLPIPTPE 618
 544 SHSND---FLTKKEGGEVPMVYANTTDKKMDLENEVYKAIINDROMPKYERYTE 600
 619 IFSFESTGTFTLYMLYKPHDLPKGYPTVLFTYGGPOVOLLVNNRFGVYFRRLNTLAS 678
 601 I-----DDVYLPQILKPAFTDTHTYPLLVVDGTGSGSVAEKFE--VSWELVWSS 652
 679 LGYVWVITDRGSGHRGKFKYKMGQIEIDQVBLQYLASRYFIDLRVGIHW 738
 653 HGAVVAVCDGSGSGFQGLLHEVRRRLGLEEKDQWEAVTML--KEGYIDRTVAVVGK 711
 739 SYGGYLSIMLMQRSD---IFRVAIGAPVTLMIFYDTGYTERYMGHPDNEQGYVLS 794
 712 DYGYLTSTYILPAKGENQGGFTFGSALSPITTDKYLASASERYLIGLHGDNAYENTK 771
 795 VAMQAEKFPSEPNLLIHGFLDENVHPARTSILSPVRAKGYDLOIYQERHSIRVP 854
 772 VAHVVSAL--EEQOFLLIHPAIDEKIHFOHTAELITOLIRKANYSIQIYDESHYFTSS 829
 855 ESGEYELHLHYQE 870
 830 SLKQHLYSITINEFVE 845
 RESULT 15
 YDZF SCHPO STANDARD: PRT; 853 AA.
 ID YDZF SCHPO
 AC Q9P7B3; O13720;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Putative dipeptidyl aminopeptidase Cl4c4.15c (EC 3.4.14.-).
 GN SPAC14c4.15c OR SPADJ760.01c
 OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryotic; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou V., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Frazer A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Kramer A.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Vancraets G., Aert R., Robben J., Grymourez B.,
RA Welljens I., Volckaert E., Rieger M., Schaefer M., Møller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Weiler H., Wambert R., Purrelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motter S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like vacuoles.
CC CC
CC -I- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AL162631; CAB83084.1; --
DR EMBL, Z98596; CAB11208.1; --
DR GeneDB, SPombe; SPAC14C4.15c; --
DR InterPro: IPR002469; DPPIV N-term.
DR InterPro: IPR001375; Peptidase S9.
DR InterPro: IPR000379; Ser-ester- site.
DR Pfam: PF00930; DPPIV N-term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
DR Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
KW Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 66 CYTOPLASMIC (POTENTIAL).
FT FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT TRANSMEM 67 89 (POTENTIAL).
FT FT
FT DOMAIN 90 853 LUMENAL (POTENTIAL).
FT ACT SITE 719 719 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 795 795 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 828 828 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 853 AA; 98341 MW; 3450DBA50F830C456 CRC64;

```

QY 156 HQSGSTFLFQAGSGYH-----VKDGGPOFTQOPLRPNLVEPSCPNIRMDPKL 204
Db 124 YHODIEWITSTGVTALVYDQSTFSLFYPDGEYGSNVDLSITTSFVLTKCKNLHR--KR 180
QY 205 CPADEWIAFIHSND-IWISN-----IYREERRLTYVHNELA-----241
Db 181 YSSDMEYIAFSCSKORRMHSHSYEDVYLVERATGRLEHLASDQSKLIVAEWSPIGHKLV 244
QY 242 -----NMEEDASAGVAIFVLQOEFDPRYS-GYWWCPKAETTPS 278
Db 241 YGIGSNLFIWESFSEPPVCITDQSDIDGLFNGNSDWVEEBELIQQSKAVWMSPDGNC--- 297
QY 279 GGGKIIRILYEENDESEVELIHY-----TSPMLFTRADSRFRPKXTGANKPTFKMSEI 333
Db 298 -----LSTLSIDDSKVP-VHVLPREQDLSKVEDQKORVNNFFHYST---FKXDPFPPVKL 346
QY 333 MIDA---EGRIDVIDKELIQPELIFEGVEYIARAGMPEGGYAMSILDRSOTRQIYV 389
Db 347 FVNCFTDSEGESLEVDSS-----FPLSTQH-RITTVAMAGNETLMVEVLRGNYERY-- 397
QY 390 LISPLFLFVEDDVNEROKRLISVPDSVTPPLIIYEETDIIWINIHIDFHV-----FPQ 442
Db 398 ---TSLF-----DLSRKTTENTREXSEHPLATSLHLKXYLSFESLGMLEKERYVROYFL 449
QY 443 SHEEELIF-----IFASECKTGF-RHLUYITSLIKESKXKRSQSGH-PAPEDF----K 489
Db 450 SNKKRIALYEEDNPVPIYITPNISFLSBLYLNNLT--YFTALSSGSEPSFRVYRLCTK 506
QY 490 CPIKEEIAITSGEWELGRHGSNIQVDEVRRLVFEYEGTQDSPLEHNLVYVSYVNPCEVTR 549
Db 507 SLILSEINIQGS---GJFGIKVANDQ-----NYLVLNVIYGP-ELPK 544
QY 550 LTRDRYS-HSCCISQCHOPFLISKYGNQKNPHCVSLYKL-----SSPEDDPCTKTEFWA 602
Db 545 --QFTYSIHEDKVTSTSDNH-SKNLPSDSSSTSGKXYLELONSJETNEBELIITTEKFA 600
QY 603 TILDSAGPLPYTPPEIFSPFESTGFTLYGMLYKHHDLOPGKKYPTVLFYGGPOYQVLV 662
Db 604 -----FSPVFVKVIKVKNITAYIQIRPNFPRKRYPRFVHLYGAPQALVT 648
QY 663 NRPFGVKYFRLNNTASLGYVVVVVIDNRGSHRGILFEGAFKXKMGQIEIDDOVEG----- 717
Db 649 GKYE---MDINELMASVYNFLVI-----KVIDVIDISD-VSQGHLES 685
QY 718 -----LOYASRYD--FIDLDRYGHMSGYGLTISLMAIMQORSIFRVAIAGAPV 765
Db 686 DSHELIITSWIELRSYVDTPYIDHHRVGIWMSGVGLTL-KILEMADPITGAVAVAV 744
QY 766 TLMIFYDTGYERYVMGHDPQONEQGYLLASVMAQAKPEPSEPRRLLLHOFLENTVAHFT 825
Db 745 TDMRYDAYVYSENLIGAVSKQTTAIYDNTVAHYSNLF-RKLCGLVLVHRTSDNVNHIENT 803
QY 826 SILSLFLVRAG---KPYDLOQIYPOEHRISIRVP 854
Db 804 MQLTKAMVEKGVNYYPR---IYVNAHNEFSDP 833

```

Search completed: October 15, 2003, 17:11:26
Job time : 20.9931 secs

Query Match 6.2%; Score 293; DB 1; Length 853;
Best Local Similarity 22.0%; Pred. No. 1.9e-13;
Matches 179; Conservative 110; Mismatches 308; Indels 216; Gaps 38;


```

Db      61 AKAPHDMEFYKRNPDGPHSDRIYYLAMSGENRENTLFYSEIPTKTINRAAVLMSKPL 120
Qy      121 DLFOATLDYGYMSSEEBELRERKRGTVGASVDYHOGSGTFLQAGSGIYHVDGGQG 180
Db      121 DLFOATLDYGYMSSEEBELRERKRGTVGASVDYHOGSGTFLQAGSGIYHVDGGQG 180
Qy      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDMIAFIHNSNDIWSNIVTREERLTYVHNL 240
Db      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDMIAFIHNSNDIWSNIVTREERLTYVHNL 240
Qy      241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEIIVH 300
Db      241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEIIVH 300
Qy      301 TSPMLERFRRADSFYPTKGTANPKVTFKMSIIMDAEGRIIDVIDKELIQFELLFEGVE 360
Db      301 TSPMLERFRRADSFYPTKGTANPKVTFKMSIIMDAEGRIIDVIDKELIQFELLFEGVE 360
Qy      361 YIARAGWTPBCKYAMSILLDRSQTRLOIVLISPFLFIVEDDVMERORLIESVDSVTP 420
Db      361 YIARAGWTPBCKYAMSILLDRSQTRLOIVLISPFLFIVEDDVMERORLIESVDSVTP 420
Qy      421 IYEBETDWINIHNDIHFVPOSHHEEIEFIASECKTGFRHLKYITSILKESYKXSSG 480
Db      421 IYEBETDWINIHNDIHFVPOSHHEEIEFIASECKTGFRHLKYITSILKESYKXSSG 480
Qy      481 GLPAPSDFKCPKEKEIATSGEWEVLGRHGSNIQVDEVRLVYEGTKDPSLEHHLVVS 540
Db      481 GLPAPSDFKCPKEKEIATSGEWEVLGRHGSNIQVDEVRLVYEGTKDPSLEHHLVVS 540
Qy      541 YVNGEVTRLTDRGYSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600
Db      541 YVNGEVTRLTDRGYSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600
Qy      601 WATILDSAGPLDPTPEIISFESTTGFTLYGMLYKPHDLOPKKYPVTLFIYGGPOVOL 660
Db      601 WATILDSAGPLDPTPEIISFESTTGFTLYGMLYKPHDLOPKKYPVTLFIYGGPOVOL 660
Qy      661 VNNRFKGVKFRINTLASLGYVVVVIDNRGSGHRGLKFEAGFKYKMGQIEIDQVEGLQY 720
Db      661 VNNRFKGVKFRINTLASLGYVVVVIDNRGSGHRGLKFEAGFKYKMGQIEIDQVEGLQY 720
Qy      721 LASRYDFIDLDVIGHMSYGYLSLMAIMORSIDIFRVALAGAVTLMIFDGTYTERYM 780
Db      721 LASRYDFIDLDVIGHMSYGYLSLMAIMORSIDIFRVALAGAVTLMIFDGTYTERYM 780
Qy      781 GHPDQNEQGYLGSVMAQAEKFPSEPRNLLIHGFLENVHFAHTSILISFLVAGKPYD 840
Db      781 GHPDQNEQGYLGSVMAQAEKFPSEPRNLLIHGFLENVHFAHTSILISFLVAGKPYD 840
Qy      841 LQIYPOERHSIRVPSGEHVELHLHYLOENLSRIAALKVI 882
Db      841 LQIYPOERHSIRVPSGEHVELHLHYLOENLSRIAALKVI 882

RESULT 2
ID      081WG7      PRELIMINARY;      PRT;      882 AA.
AC      081WG7;
DT      01-MAR-2003 (Tremblrel. 23, Created)
DT      01-MAR-2003 (Tremblrel. 23, last sequence update)
DT      01-MAR-2003 (Tremblrel. 23, last annotation update)
DE      Similar to dipeptidylpeptidase 8.
OS      Homo sapiens (Human).
OC      Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Strausberg R.;
RL      Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.

```

```

DR      EMBL; BC040203; AA040203.1; -.
SQ      SEQUENCE      882 AA; 101391 MM; 88C76A5BCE707F9 CRC64;
Query Match      99.9%; Score 4696; DB 4; Length 882;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 881; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy      1 MAAMTEQUGVEIFETADCEENIESODRPLEFFYERYYSWSQJLKLADTRYKHGYM 60
Db      1 MAAMTEQUGVEIFETADCEENIESODRPLEFFYERYYSWSQJLKLADTRYKHGYM 60
Qy      61 AKAPHDMEFYKRNPDGPHSDRIYYLAMSGENRENTLFYSEIPTKTINRAAVLMSKPL 120
Db      61 AKAPHDMEFYKRNPDGPHSDRIYYLAMSGENRENTLFYSEIPTKTINRAAVLMSKPL 120
Qy      121 DLFOATLDYGYMSSEEBELRERKRGTVGASVDYHOGSGTFLQAGSGIYHVDGGQG 180
Db      121 DLFOATLDYGYMSSEEBELRERKRGTVGASVDYHOGSGTFLQAGSGIYHVDGGQG 180
Qy      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDMIAFIHNSNDIWSNIVTREERLTYVHNL 240
Db      181 FTQOPLRPNLVETSCPNIRMDPKLCPADPDMIAFIHNSNDIWSNIVTREERLTYVHNL 240
Qy      241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEIIVH 300
Db      241 ANMEDARSAGVATFVLOEEDRYSGYWMCCKAETTPSGGKILRIIYEENDESEVEIIVH 300
Qy      301 TSPMLERFRRADSFYPTKGTANPKVTFKMSIIMDAEGRIIDVIDKELIQFELLFEGVE 360
Db      301 TSPMLERFRRADSFYPTKGTANPKVTFKMSIIMDAEGRIIDVIDKELIQFELLFEGVE 360
Qy      361 YIARAGWTPBCKYAMSILLDRSQTRLOIVLISPFLFIVEDDVMERORLIESVDSVTP 420
Db      361 YIARAGWTPBCKYAMSILLDRSQTRLOIVLISPFLFIVEDDVMERORLIESVDSVTP 420
Qy      421 IYEBETDWINIHNDIHFVPOSHHEEIEFIASECKTGFRHLKYITSILKESYKXSSG 480
Db      421 IYEBETDWINIHNDIHFVPOSHHEEIEFIASECKTGFRHLKYITSILKESYKXSSG 480
Qy      481 GLPAPSDFKCPKEKEIATSGEWEVLGRHGSNIQVDEVRLVYEGTKDPSLEHHLVVS 540
Db      481 GLPAPSDFKCPKEKEIATSGEWEVLGRHGSNIQVDEVRLVYEGTKDPSLEHHLVVS 540
Qy      541 YVNGEVTRLTDRGYSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600
Db      541 YVNGEVTRLTDRGYSHSCCISOHCDFFISKYSNOKNPHCVSLYKLSPEDDPTCKTKEF 600
Qy      601 WATILDSAGPLDPTPEIISFESTTGFTLYGMLYKPHDLOPKKYPVTLFIYGGPOVOL 660
Db      601 WATILDSAGPLDPTPEIISFESTTGFTLYGMLYKPHDLOPKKYPVTLFIYGGPOVOL 660
Qy      661 VNNRFKGVKFRINTLASLGYVVVVIDNRGSGHRGLKFEAGFKYKMGQIEIDQVEGLQY 720
Db      661 VNNRFKGVKFRINTLASLGYVVVVIDNRGSGHRGLKFEAGFKYKMGQIEIDQVEGLQY 720
Qy      721 LASRYDFIDLDVIGHMSYGYLSLMAIMORSIDIFRVALAGAVTLMIFDGTYTERYM 780
Db      721 LASRYDFIDLDVIGHMSYGYLSLMAIMORSIDIFRVALAGAVTLMIFDGTYTERYM 780
Qy      781 GHPDQNEQGYLGSVMAQAEKFPSEPRNLLIHGFLENVHFAHTSILISFLVAGKPYD 840
Db      781 GHPDQNEQGYLGSVMAQAEKFPSEPRNLLIHGFLENVHFAHTSILISFLVAGKPYD 840
Qy      841 LQIYPOERHSIRVPSGEHVELHLHYLOENLSRIAALKVI 882
Db      841 LQIYPOERHSIRVPSGEHVELHLHYLOENLSRIAALKVI 882

RESULT 3
ID      09D4G6      PRELIMINARY;      PRT;      892 AA.
AC      09D4G6;
DT      01-JUN-2001 (Tremblrel. 17, Created)

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Adult male testis cDNA, RIKEN full-length enriched library,
 DE clone:493244F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteleia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RA Aachai J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arawaka T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoaka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Saeki D., Shibata K., Shibata Y., Shingawa A., Shitaka T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=22354683; PubMed=1246851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.,
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=2049374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kasahara K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe K.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunaga S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 DR EMBL, AK016546; BAB30295.2;
 SQ SEQUENCE 892 AA; 102284 MM; FDE0DBDDC4CA346 CRC64;
 Query Match 96.4%; Score 4528.5; DB 11; Length 892;

Best Local Similarity 95.7%; Pred. No. 0;
 Matches 845; Conservative 18; Mismatches 19; Indels 1; Gaps 1;
 QY 1 MAAMETBQGVLEIFETADCE-NISQDQPKLEPFVVEYSSQKLLADPRKHGYM 59
 DB 10 MAAMETBQGVLEIFETADCEBNGSODPKLEPFVVEYSSQKLLADPRKHGYM 69
 QY 60 MAKAPHDPMFKNDPQGSDBRIYYLAMSGENRENTLFYSEIPIKTIINRAAVLMSKPL 119
 DB 70 MAKAPHDPMFKNDPQGSDBRIYYLAMSGENRENTLFYSEIPIKTIINRAAVLMSKPL 129
 QY 120 LDFPQATLDYGMTSREBELREKRIGTGTGIAVDYHQSQTFLPQASGIVHYKDGQP 179
 DB 130 LDFPQATLDYGMTSREBELREKRIGTGTGIAVDYHQSQTFLPQASGIVHYKDGQP 189
 QY 180 GFTQOPRLPVLTSCNINIMDKLCPADDDWIAFHSNINIMSNITRERRRLTYVNE 239
 DB 190 GFTQOPRLPVLTSCNINIMDKLCPADDDWIAFHSNINIMSNITRERRRLTYVNE 249
 QY 240 LANNEEDARSAGVATFVLOREDFRYSQYMWCPAETTPSGKILRIYEENDESEVEI 299
 DB 250 LANNEEDARSAGVATFVLOREDFRYSQYMWCPAETTPSGKILRIYEENDESEVEI 309
 QY 300 VTSPMLSTRADSPRYPKTGTANPKYTFKASEIMIDAGRIIDYIDKELIQPEFILFEGV 359
 DB 310 VTSPMLSTRADSPRYPKTGTANPKYTFKASEIMIDAGRIIDYIDKELIQPEFILFEGV 369
 QY 360 EYIARAGWTPREGVYASILLDRSOTRLQVILISPELFIPEDDMMERQRLIESVPDSTP 419
 DB 370 EYIARAGWTPREGVYASILLDRSOTRLQVILISPELFIPEDDMMERQRLIESVPDSTP 429
 QY 420 LIYEETTDIMINIHDIHFVFPQSHRELEIFASECKTFRHLKYKTSILKSKYRSS 479
 DB 430 LIYEETTDIMINIHDIHFVFPQSHRELEIFASECKTFRHLKYKTSILKSKYRSS 489
 QY 480 GGIAPASDFPCPIKEEALITSGEWVLGRGNSINQVDEVRLTYFEGTKOSPLEHLLYV 539
 DB 490 GGIAPASDFPCPIKEEALITSGEWVLGRGNSINQVDEVRLTYFEGTKOSPLEHLLYV 549
 QY 540 SYVNGPVPRLTGRGSHSCCISOHCDFFISXSNQNCPCVSLYKLSPEDDPTCKTKE 599
 DB 550 SYVNGPVPRLTGRGSHSCCISOHCDFFISXSNQNCPCVSLYKLSPEDDPTCKTKE 609
 QY 600 FMATILDSAGPLPDYTPPELIFSPESSTGFTLYGMLYKPHDLOPKKYPVLFYIGSPQV 659
 DB 610 FMATILDSAGPLPDYTPPELIFSPESSTGFTLYGMLYKPHDLOPKKYPVLFYIGSPQV 669
 QY 660 LVNNRFKGYKYPFLNLTASLGIYVVYVINDRGSCHRGKKEFGAKRYKMGQIEIDQVEGLQ 719
 DB 670 LVNNRFKGYKYPFLNLTASLGIYVVYVINDRGSCHRGKKEFGAKRYKMGQIEIDQVEGLQ 729
 QY 720 YLASRYDFIDLDEVGTHGMSYGYLSLMLMORSDFRVAIAGAPVTLMTFYDTGTERY 779
 DB 730 YLASRYDFIDLDEVGTHGMSYGYLSLMLMORSDFRVAIAGAPVTLMTFYDTGTERY 789
 QY 780 MGHPDONEQGYLYGSVAMQAEKPESEPNRLLLHGFIDENVHFAHTSILSLFVRAKPY 839
 DB 790 MGHPDONEQGYLYGSVAMQAEKPESEPNRLLLHGFIDENVHFAHTSILSLFVRAKPY 849
 QY 840 DLQIYPOBHSITVPESGHEHLHLHYOENIGSRIALXV 882
 DB 850 DLQIYPOBHSITVPESGHEHLHLHYOENIGSRIALXV 892
 RESULT 4
 O8NEMS PRELIMINARY; PRT; 831 AA.
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Similar to dipeptidyl peptidase 8.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Strausberg R.;
 RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC030688; AAH30688.1;
 SQ SEQUENCE 831 AA; 95527 MW; 0B2A13A2FE70CB2 CRC64;

Query Match 93.6%; Score 4397.5; DB 4; Length 831;
 Best Local Similarity 94.2%; Pred. No. 0;
 Matches 831; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

1 MAAMETEQGLVETFEADCEENIESODRPLEPFYERYYSQTLKILATRRYHYGM 60
 1 MAAMETEQGLVETFEADCEENIESODRPLEPFYERYYSQTLKILATRRYHYGM 60
 61 AKAPHEMFVKNRDPDPHSDRIYYLAMSNGENRENTLFYSPIKTINRAAVLMSKPL 120
 61 AKAPHEMFVKNRDPDPHSDRIYYLAMSNGENRENTLFYSPIKTINRAAVLMSKPL 120
 121 DLPAATLDYGMYSSEELPERKIGTVGTAHYHOGSGTFLFOAGSGIYHVDGPG 180
 121 DLPAATLDYGMYSSEELPERKIGTVGTAHYHOGSGTFLFOAGSGIYHVDGPG 180
 121 DLPATLDYGMYSSEELPERKIGTVGTAHYHOGSGTFLFOAGSGIYHVDGPG 180
 181 FTQGPLRPNIYETSCPNIRMDPKLCPADPWIAFIHNSNDIWSNIVTREERLLTYANEL 240
 181 FTQGPLRPNIYETSCPNIRMDPKLCPADPWIAFIHNSNDIWSNIVTREERLLTYANEL 240
 241 ANMEDARSAGVATFVYQEEFDRISGYWCPKATTPSGGKILRIYENDSEVEIIVH 300
 241 ANMEDARSAGVATFVYQEEFDRISGYWCPKATTPSGGKILRIYENDSEVEIIVH 300
 241 ANMEDARSAGVATFVYQEEFDRISGYWCPKATTPSGGKILRIYENDSEVEIIVH 300
 301 TSPMLERRRDSFPYKGTANPKYTFMSEIMIDAGRIIDVDKELIQPELIFEGVE 360
 301 TSPMLERRRDSFPYKGTANPKYTFMSEIMIDAGRIIDVDKELIQPELIFEGVE 360
 301 TSPMLERRRDSFPYKGTANPKYTFMSEIMIDAGRIIDVDKELIQPELIFEGVE 360
 361 YIARAGWTPEGKYAMSILLDRSQTRLQIVLISPELFIPEDDVWERORLIESVDSVTP 420
 361 YIARAGWTPEGKYAMSILLDRSQTRLQIVLISPELFIPEDDVWERORLIESVDSVTP 420
 421 IYEEETDWINIHDIHIVPQSHHEBIEFIASECTGFRHLKYTTSIIKESKYRSSG 480
 421 IYEEETDWINIHDIHIVPQSHHEBIEFIASECTGFRHLKYTTSIIKESKYRSSG 480
 421 IYEEETDWINIHDIHIVPQSHHEBIEFIASECTGFRHLKYTTSIIKESKYRSSG 480
 481 GLPAPSPFCPIKEEIAITSGEWEVLGRHGSNIQVDEVRILYVEGTKDSPLRHLIYVS 540
 481 GLPAPSPFCPIKEEIAITSGEWEVLGRHGSNIQVDEVRILYVEGTKDSPLRHLIYVS 540
 481 GLPAPSPFCPIKEEIAITSGEWEVLGRHGSNIQVDEVRILYVEGTKDSPLRHLIYVS 540
 541 YVNPGEVTRLDRGYSHSCISQHCDFPISKYSNOKNPHCVSLYKLSPPEDDPTCKTEF 600
 541 YVNPGEVTRLDRGYSHSCISQHCDFPISKYSNOKNPHCVSLYKLSPPEDDPTCKTEF 600
 541 YVNPGEVTRLDRGYSHSCISQHCDFPISKYSNOKNPHCVSLYKLSPPEDDPTCKTEF 600
 601 MATLLDSAGLPLDTPPEIISFESTTFTLYGMLYKHDIQPGKYTVVFIYGGPOVL 660
 601 MATLLDSAGLPLDTPPEIISFESTTFTLYGMLYKHDIQPGKYTVVFIYGGPOVL 660
 601 MATLLDSAGLPLDTPPEIISFESTTFTLYGMLYKHDIQPGKYTVVFIYGGPOVL 660
 661 VNNRFGKGYKFRLLNTLASLGYVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGL 720
 661 VNNRFGKGYKFRLLNTLASLGYVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGL 720
 661 VNNRFGKGYKFRLLNTLASLGYVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGL 720
 721 VNNRFGKGYKFRLLNTLASLGYVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGL 720
 721 VNNRFGKGYKFRLLNTLASLGYVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGL 720
 721 VNNRFGKGYKFRLLNTLASLGYVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGL 720
 707 VNNRFGKGYKFRLLNTLASLGYVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGL 720
 707 VNNRFGKGYKFRLLNTLASLGYVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGL 720
 707 VNNRFGKGYKFRLLNTLASLGYVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGL 720
 781 GHPDNOEGYYLGSVAMQAEKPPSEPRNLLHGFLENNVHFAHTSILSLVRAAGPYD 840
 781 GHPDNOEGYYLGSVAMQAEKPPSEPRNLLHGFLENNVHFAHTSILSLVRAAGPYD 840
 781 GHPDNOEGYYLGSVAMQAEKPPSEPRNLLHGFLENNVHFAHTSILSLVRAAGPYD 840
 841 LQIYPERHSIRVPESGEHYELHLHYQENLGSRIALAKVI 882
 841 LQIYPERHSIRVPESGEHYELHLHYQENLGSRIALAKVI 882
 841 LQIYPERHSIRVPESGEHYELHLHYQENLGSRIALAKVI 882

Db 790 LQIYPERHSIRVPESGEHYELHLHYQENLGSRIALAKVI 831

RESULT 5

096JX1 PRELIMINARY; PRT; 632 AA.
 ID 096JX1;
 AC 096JX1;
 DT 01-DEC-2001 (TREMUREL 19, Created)
 DT 01-DEC-2001 (TREMUREL 19, Last sequence update)
 DT 01-OCT-2002 (TREMUREL 22, Last annotation update)
 DE Hypothetical protein FLJ14920.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
 RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Takiguchi S.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Ishii S., Kawai Y., Saito K.,
 RA Niimura K., Iwayanagi T.;
 RT "NBD human cDNA sequencing project."
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK027426; BAB55395.1;
 DR MEROPS; S09_018;
 KW Hypothetical protein.
 SQ SEQUENCE 632 AA; 72639 MW; 9BDF598B06985AA4 CRC64;

Query Match 70.9%; Score 333.5; DB 4; Length 632;
 Best Local Similarity 92.4%; Pred. No. 1.9e-250;
 Matches 631; Conservative 0; Mismatches 1; Indels 51; Gaps 1;

200 MDKELCPADPDWIAFIHNSNDIWSNIVTREERLLTYVHNEIAMEEDARSAGVATFVLOE 259
 1 MDKELCPADPDWIAFIHNSNDIWSNIVTREERLLTYVHNEIAMEEDARSAGVATFVLOE 259
 260 EPRYSYGYWCPKATTPSGGKILRIYENDSEVEIIVHVSPTMLETTRADSPFRYKGT 319
 61 EPRYSYGYWCPKATTPSGGKILRIYENDSEVEIIVHVSPTMLETTRADSPFRYKGT 319
 320 TANPKYTFKMSIEMIDAGRIIDVDKELIQPELIFEGVEYIARAGWTPEGKYAMSILL 379
 121 TANPKYTFKMSIEMIDAGRIIDVDKELIQPELIFEGVEYIARAGWTPEGKYAMSILL 379
 380 DRSQTRLQIVLISPELFIPEDDVWERORLIESVDSVTPLIYEEETDWINIHDIHIV 439
 181 DRSQTRLQIVLISPELFIPEDDVWERORLIESVDSVTPLIYEEETDWINIHDIHIV 439
 440 FPGSHHEEIEFIASECTGFRHLKYTTSIIKESKYRSSGGLPAPSPFCPIKEEIAIT 499
 241 FPGSHHEEIEFIASECTGFRHLKYTTSIIKESKYRSSGGLPAPSPFCPIKEEIAIT 499
 500 SGEWEVLGRHGSNIQVDEVRILYVEGTKDSPLRHLIYVSYPNPGEVTRLDRGYSHSC 559
 301 SGEWEVLGRHGSNIQVDEVRILYVEGTKDSPLRHLIYVSYPNPGEVTRLDRGYSHSC 559
 560 CISOHCDFPISKYSNOKNPHCVSLYKLSPPEDDPTCKTEFMAITLDSAGPLDTPPEI 619
 361 CISOHCDFPISKYSNOKNPHCVSLYKLSPPEDDPTCKTEFMAITLDSAGPLDTPPEI 619
 620 FPFESTTFTLYGMLYKHDIQPGKYTVVFIYGGPOVLVNNRFGKGYKFRLLNTLASL 679
 421 FPFESTTFTLYGMLYKHDIQPGKYTVVFIYGGPOVLVNNRFGKGYKFRLLNTLASL 679
 680 GYVVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGLYASRYDFIDLDVGIHGS 739
 481 GYVVVVINDRGSCHRGKLFEGAFKYMGOIETDDQVEGLYASRYDFIDLDVGIHGS 739
 740 YGYSLSLMLMORSDFRVAIAGAPVTLMIFYDTGTYTTERYMGHPDNOEGYYLGSVAMQ 799
 740 YGYSLSLMLMORSDFRVAIAGAPVTLMIFYDTGTYTTERYMGHPDNOEGYYLGSVAMQ 799
 740 YGYSLSLMLMORSDFRVAIAGAPVTLMIFYDTGTYTTERYMGHPDNOEGYYLGSVAMQ 799

```

Db 508 -----VALAGAPVIMITYDGYTERYMGHPDQNGGYLSVAAQA 549
Qy 800 EKFPSEPNRLLLHGFIDENVFAHTSILLSFIVAGKPYDLOIYPOHRHSIRVESGEH 859
Db 550 EKFPSEPNRLLLHGFIDENVFAHTSILLSFIVAGKPYDLOIYPOHRHSIRVESGEH 609
Qy 860 YELHLHYLOENLGSRIALAVYI 882
Db 610 YELHLHYLOENLGSRIALAVYI 632

RESULT 6
Q8BXD8 PRELIMINARY; PRT; 863 AA.
ID Q8BXD8
AC Q8BXD8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Dipeptidyl peptidase 9.
GN DPP9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1];
RF SEQUENCE FROM N.A.
RA Olsen C., Maglmann N.;
RT "Identification and characterization of a novel member of the
RT dipeptidyl peptidase IV-related family";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [2];
RF SEQUENCE FROM N.A.
RA Strauberg R.;
RT Tissue=Skin;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF452102; AAL47179.1; -.
DR EMBL; BC037948; AAH37948.1; -.
DR MEROPS; S09.019; -.
DR InterPro; IPR002469; DPPIV N term.
DR InterPro; IPR001375; Peptidase S9.
DR InterPro; IPR000379; Ser esterase site.
DR Pfam; PF00930; DPPIV N term; 1.
DR Pfam; PF00326; Peptidase S9; 1.
SQ SEQUENCE 863 AA; 98263 MW; 40FEDB78E26CDED5 CRC64;

Query Match 61.1%; Score 2870; DB 4; Length 863;
Best Local Similarity 61.5%; Pred. No. 3.4e-214;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;
Qy 35 FYERYWSQKLKLLADTRKYHYMAKAPHDPMFVKRNDPDGHSRIYYLAMSNGRE 94
Db 24 FOVKHSDWGLRSIIHSGSRKSGILVINKAPHDFOVKTESGPHSHRYLLGMPYSRE 83
Qy 95 NTFYSEIPTINRAAVLMSKRLDLPQATLDYGYMSREBELRERKIGTVGIASVD 154
Db 84 NSLIYSIIPKVKRKAELLLSWKQMDLHFQATPHGYVSREELRERKLGVGITSYD 143
Qy 155 YHOGSGTFLEOAGSGIYHVDGPGFTQOPLRPNLVETSCPNIRMPKICPADPDIAT 214
Db 144 FHESBSGLFLFOASNSLPHCHDGGKNGWSPMKLEIKTQCSGRMDPKICPADPAFSP 203
Qy 215 IHSNDIWSINIVREERLLTYVANELANMEEDASAGAVATFVIOEFDKYSGVWCPKAE 274
Db 204 INNSDMVANIETGEERRLTFCHQGLSNVLDPPKASAGVATFVIOEFDKFTGYWCPKAS 263
Qy 275 TTPSGG-KILRIIYENDESEVELIHYTSPMLFTRRADSPRYPTGTANRKYVFKSEIM 333
Db 264 WBSBSGLKTRIRIYEVDESEVEYIHPSPALBERKTDSTRYPTGSKNFKIAIKLAEFQ 323
Qy 334 IDAEGRIIDVDELQIOPFELFEGVYIARAGWTPRGKAYMSILLDRSOTRLQIVLISP 393
Db 324 TDSQGIKIVTQEKELVQPFSSLPKVKYIARAGWTRDGKAYAMMFLDRPQOWIQIVLILP 383

```

```

Qy 394 ELFIPEVDYMERORLLIESVDSYTPLLIYEETDIWINIMIDIRHPVQSH-EELIERIF 452
Db 384 ALFIPESTENEBQLASARAAPRVNQPYYVEEVTNWINVMDIFYPFQSGEBDELCTLR 443
Qy 453 ASECKTGRRLYKITSILKESKYKSSGGLPAPSDFKPIKEEIALISGEWVYGRHSN 512
Db 444 ANECKTGFCHLYKTAVLKSQYDMSERPFSGEDEFKPIKEEIALISGEWVYARHSK 503
Qy 513 IQDEVRILYFEGCKDSPLEHNLVYVSYNVRGEVTRITLDGSHSCCISQHCDFISKY 572
Db 504 IYVNEETLVYFQGTCTPLLEHLHYVVSYEAGEIVRLITGFSHSCMSQNFDMFVSHY 563
Qy 573 SNQKPRCVSLYKSSDEDDPTCKTEFWATILDSAGLPDYTPPEIRFSSTQFTLYG 632
Db 564 SSVSTPEVHYKILSGDDDDLHQPFRFASMEASCPDYVPEIRIHFHTRSVRLYG 623
Qy 633 MLYRPHLOPKKPYTYLFTYGGPOVOLVNNRFGVKYFRINTASLGYVVVVDNRSGC 692
Db 624 MIYRPHALOPKPKPTVLFYVGGPOVOLVNNRFGIKYLRINTASLGYAVVVIDGRSGC 683
Qy 693 HRGLKFGAFKYGKGOIEIDDOVEGLQYLASRYDFIDLRGIGHGWSYGYLSIMALMOR 752
Db 684 QRGLEFGALKNQGOVEIEDQVEGLQPVAKYGFIDLSRVALHGWSTGYGLSLMGLTHK 743
Qy 753 SDIFRVAIAGAPVTLMTFYDTGYTERYMGHPDQNGGYLGSVAAQAEKFPSEPNRLLL 812
Db 744 PQVFKAIALAGAPVTVMMAVDGYTERYMDVDENNQHGEVAGSVALHVEKLNENPRLLIL 803
Qy 813 HGFIDENVFAHTSILLSFIVAGKPYDLOIYPOHRHSIRVESGEHYELHLHYLOENL 872
Db 804 HGFIDENVFAHTSILLSFIVAGKPYDLOIYPOHRHSIRVESGEHYEVTLLHFLQYTL 863

RESULT 7
Q8BWT9 PRELIMINARY; PRT; 862 AA.
ID Q8BWT9
AC Q8BWT9;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Dipeptidyl peptidase 9 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1];
RF SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002).
DR EMBL; AK050021; BAC34034.1; -.
SQ SEQUENCE 862 AA; 98001 MW; B1D56E824A834B8 CRC64;

Query Match 60.3%; Score 2833; DB 11; Length 862;
Best Local Similarity 60.8%; Pred. No. 2.6e-211;
Matches 511; Conservative 134; Mismatches 193; Indels 2; Gaps 2;
Qy 35 FYERYWSQKLKLLADTRKYHYMAKAPHDPMFVKRNDPDGHSRIYYLAMSNGRE 94
Db 23 FOVKHSDWGLRSIIHSGSRKSGILYSAKADHQPQVKPESGPHSHRYLLGMPYSRE 82
Qy 95 NTFYSEIPTINRAAVLMSKRLDLPQATLDYGYMSREBELRERKIGTVGIASVD 154
Db 83 NSLIYSIIPKVKRKAELLLSWKQMDLHFQATPHGYVSREELRERKLGVGITSYD 142
Qy 155 YHOGSGTFLEOAGSGIYHVDGPGFTQOPLRPNLVETSCPNIRMPKICPADPDIAT 214
Db 143 FHESBSGLFLFOASNSLPHCHDGGKNGWSPMKLEIKTQCSGRMDPKICPADPAFSP 202

```

QY	215	HSN01WISNIYTRERRRLTYVHNELJANEEDRASGVATFVLYQEEFDDISGVWCKPKAE	274
Db	203	INNSDLWMANIETGEERRLITFCQSGAGVLDNKSAGVATFVLYQEEFDDITGCMGCPYAS	262
QY	275	TTPSGG-KILRLIYEENDESEVEI IHTVSPMLETRADSFRYPKGTANPKVYFKMSEIM	333
QY	334	IDAEERIIDVIDKEILOPELLEPEGVEYIARACMTPEGKYAMSILIDRSOTRLQIYLISB	393
Db	323	TDHOGKYSSCKEELVQPPSSLFPEKVEYIARACMTDGGYAMAMFIDRQOORLOIYLLEP	382
QY	394	ELFIEVEDDVMERORLIASVPDVTPLIYEETDITWIMIHDAIFHYPOSH--EEIYEFIF	452
Db	383	ALFIFAVESEAQROAAARAVPKNVQPFVYIEEVTWIMIVHOLFHFPPAQEGQDCFLR	442
QY	453	ASECTGFPHLYKITSILKESKRYKSSGGLPADSDFCPIKEBIAITSGEMVLYGHGNSN	512
Db	443	ANBECTGFCHLYRVVVELTKDYDWTPELPSPEDEKCPIKEBVALTSGEMEVLSPHGSK	502
QY	513	IQVDERRLRYVEEGTKDPSLEHNLVYVSVNPEDEVRLIDRGVSHSCCISQHDPFISXY	572
Db	503	IWMNQTKLYVYQGTIKDTLEHNLVYVSESAGEIYRLITLTFGSHSCMSQSDNMFVSH	562
QY	573	SNQKNPHCVSLYKLSPEPDDPTCKTEPMATLIDSGAPLPDYTPPEIFSPFESTGFTLYG	632
Db	563	SSVSTPPCVHYVYKLSGPPDDPLPHKORFMAWSMMEANCPDDYVPEPIFFHTRADVQLYG	622
QY	633	MLYKPHDLOPGKKYPTVLYIGSGPOVLNNRFKGVKYRLMTIASLGVVVVYIDRGSC	692
Db	623	MYKHHITLOPGKHPFLVLYIGSGPOVLNNSEFKGIKYRLMTIASLGAVVVIDRGSC	682
QY	693	HRLKEBEGAFKYKMQQIEIDDOVEGLOYLASHRYDFIDLDVGVIGHGWSYGGYLSIMALMR	752
Db	683	QRLGFEGALKNMQQVEIEDQVEGLOLYAEXYGFIDLSRVALHGMWSYGGYLSIMGLHMK	742
QY	753	SDIFPVATAGAPVTLMIPIYDQGYTERYMGHPRONEGGYLISGVANQAEKFPSEPNRLIL	812
Db	743	POVFVVALAGAVVYVMAADVTGYTERYMDVPENNQGYEAGSVVALHVELPNEPNRLIL	802
QY	813	HGFLDENYHFAHTSILLSEFLVAGKRYDLOIYPOEHSIRVSESGHYELATHLYLOEYL	872
Db	803	HGFLDENYHFFITNLVSQLIRAKGRYDLOIYPRNHSIRCRSESGHYEVTLLHFLQEHIL	862
RESULT 8			
Q8BVG4	Q8BVG4	PRELIMINARY;	PRT; 862 AA.
AC	Q8BVG4		
DT	01-MAR-2003 (Tremblrel. 23, Created)		
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DS	Dipeptidyl peptidase 9 homolog.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCHI_Taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Olfactory brain;		
RX	MEDLINE=22354683; PubMed=12466851;		
RA	The FANTOM Consortium,		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		
RT	"Analysis of the mouse transcriptome based on functional annotation of		
RL	60,770 full-length cDNAs."		
DR	Nature 420:563-573(2002).		
QC	EMBL; AK078301; BAC37211.1; --		
Q8	SEQUENCE 862 AA; 98049 MW; B4FD3476B6F50030 CRC64;		

Query Match	60.1%;	Score 2823;	DB 11;	Length 862;
Best Local Similarity	60.7%;	Pred. No. 1.5e-210;		
Matches 510;	Conservative 134;	Mismatches 194;	Indels 2;	Gaps 2

QY	35	FLVEVYSSQKLKLLADTRKTHGYMAKAPDPMVYKENDPDGPHSDRIYYLAMSSEN	94
Db	23	FCVGHSHWDGJRSIIHGSRKSSGLIVSKAPHDFOFVOKDESGPHSHRYLYLGMPIYSR <td>82</td>	82
QY	95	NLFVSEI.PKTIINRAAVLMSKPLLDIFQATLDJGMSREBELRERRIGVGYASD <td>154</td>	154
Db	83	NSLTYSEL.PKKYRKALLLSLSKQMLDHFQATPHHGVNSREBELRERKLGVFGITSD <td>142</td>	142
QY	155	YHQSGETL.FQAGSGIYHVKGQPGPGFTQOPLRPNLVETSCPNIRMDPKLCRADPMIAF <td>214</td>	214
Db	143	FHSESGTL.FQASNSLPHRODGKNGKPMVSPKPLEIKTQCQGGPRMDPKICPADPAFISF <td>202</td>	202
QY	215	IHSNDIMV.SNITYTBERRLTYVHNELANMEDARSAGVALTYLQEBFDYSGYWCCKAE <td>274</td>	274
Db	203	INNSLTMVANIETGERRLTFCHQSGAGVLDPKAGAVATFYIQEEFDFTGCWCPITAS <td>262</td>	262
QY	275	TTPSGG-ILIRLYENDESEVEIITHVSPMLSTRADSPR.PKXTANPKYTFKSEIM <td>333</td>	333
Db	263	WEGSGGLKTLRLVHSEVSEVEVTHVSPALBERKTDYR.PRYGTSKNPKIALKALAEIO <td>322</td>	322
QY	334	IDABGRILDIVDKELIQEFELIFEEVEYIARAGWPBECKYAMSILDRSOTRLQVILSP <td>393</td>	393
Db	323	TDHQKIVSSCEKELVQPFSSLF.PKVEYIARAGWTRDGKYAMAFLYRQORQLVLLRP <td>382</td>	382
QY	394	ELFIFVEDWNERQGLIESVPSTPLIIEETDITMINHDIHFVFPQSH-EEELFELF <td>452</td>	452
Db	383	ALFIPAVSEARQARAPVKNQVPFVIEEVTVMVNHDIHFHPDABEGQDDCFILR <td>442</td>	442
QY	453	ASECTGPRHLYKITSILKESKYKSSGGLPAPSPKPCIKELIATSGEWELGHHGS <td>512</td>	512
Db	443	ANECKTGCHLYRVTELKTDYMTPELSPDEDFKCIKEVALTSGEWEVLSHRGSK <td>502</td>	502
QY	513	IQVDEVRLLVFEETKDSPLHHLLVVSVYVNGEYTRLLDRGYSHSCCISQHCDFISKY <td>572</td>	572
Db	503	IMVNBQTKLVYFGQTKDTPLEHHLVVSYESAGEIVRLITLGLFHSQCSMSQSPDMVSHY <td>562</td>	562
QY	573	SNQKRPCHVSLYKLSSEPDDBPCKTKEWATITLSAGPLPDYTPPELISFESBTGTYLG <td>632</td>	632
Db	563	SSVSTPCCVHYKJLSGPDDBPLHKRPRFMASSMEANCPDPVPEIIFHHTRADVQLTG <td>622</td>	622
QY	633	MLYPRHDQOPKKRPTVFIYIGGPOVOLVNNRFKVCYFRNLTLASLGVVVVINDRGSC <td>692</td>	692
Db	623	MIYPRHTIQPKRHPTVFVYGGPOVOLVNNSPKSIKYLRLNTLTLASLGVAVVVDGRGSC <td>682</td>	682
QY	693	HRLKPEGAFKYKMGQLEIDDQVBSGLYLAARYDFILDRYGIHMSYGYLSLTMALMOR <td>752</td>	752
Db	683	QRLHFEGLAKMGQVIEIDQVEGLQVAAEKYGRITLSRAVHMSYGGFLSLMGLIHK <td>742</td>	742
QY	753	SDIRVALAAGPVLMTFYDTGYTBYRYVGHDPDQEGYLLSGVAMQAEKFPSEPNRLIL <td>812</td>	812
Db	743	POVERVALAAGPVTVMAYVDTGYTYRYVDPVNNKQGEAGSVAAVHEL.PNEPRRLTL <td>802</td>	802
QY	813	HGFIDENVHFAHTSILSFLVRAGKPYLQIYPOGRHSIRVPESGEHMLHTLYLOENT <td>872</td>	872
Db	803	HGFIDENVHFTHTFVLSQILRAGKPYLQIYPERHSIRCRESGEHEVTLTHFLQEHNL <td>862</td>	862
RESULT 9			
Q9NXP4 PRELIMINARY; PRT; 469 AA.			
AC 01-OCT-2000 (TREMBLrel. 15, Created)			
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)			
DE 01-OCT-2002 (TREMBLrel. 22, last annotation update)			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,			

RA NCBI_taxid=9606;
RN [1]
SEQUENCE FROM N.A.
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,

RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
 RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isegawa T., Sugano S.,
 RT "NEDO human cDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AK000230; BAA91059.1; -
 DR MEROPS; S09.018; -
 KW Hypothetical protein.

SO SEQUENCE 469 AA; 54367 MW; 088ED0B1E46C11F CRC64;

Query Match 52.8%; Score 2483; DB 4; Length 469;
 Best Local Similarity 100.0%; Pred. No. 1,7e-184;
 Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAMETEQGLVEIFETADCEENIESQDPKLEPPYERYSWQLKTLADTRYKGYMM 60
 DB 1 MAAMETEQGLVEIFETADCEENIESQDPKLEPPYERYSWQLKTLADTRYKGYMM 60
 QY 61 AKAPHDPMFVKRNDPDGPHSDRIYLLAMSGENRENTLFYSRIPTINRAALMSWPL 120
 DB 61 AKAPHDPMFVKRNDPDGPHSDRIYLLAMSGENRENTLFYSRIPTINRAALMSWPL 120
 QY 121 DLFOATLDYGYMSSEBELRERKRGTVGLASVDYHOGSGTFLFOAGSGIYHVKGPG 180
 DB 121 DLFOATLDYGYMSSEBELRERKRGTVGLASVDYHOGSGTFLFOAGSGIYHVKGPG 180
 QY 181 FTQOPARPNLVESCPYIMDPKCPADPMIAFHSNDTMSINVTREERLLTYVNEL 240
 DB 181 FTQOPARPNLVESCPYIMDPKCPADPMIAFHSNDTMSINVTREERLLTYVNEL 240
 QY 241 ANNEEDARSAGVATFVQSEEDRYSGYWMCFAETTPSGKILRLYEENDESEVEIHV 300
 DB 241 ANNEEDARSAGVATFVQSEEDRYSGYWMCFAETTPSGKILRLYEENDESEVEIHV 300
 QY 301 TSPMLERRADSEFPYPTGTANPKVTFKMSIMIDAGRILIDVIXELIQFELLFEGVE 360
 DB 301 TSPMLERRADSEFPYPTGTANPKVTFKMSIMIDAGRILIDVIXELIQFELLFEGVE 360
 QY 361 YIIRAGWTPGKAYMSILLRSQRLQIVLISPLFIPVEDDWERQRLIESVDSVPL 420
 DB 361 YIIRAGWTPGKAYMSILLRSQRLQIVLISPLFIPVEDDWERQRLIESVDSVPL 420
 QY 421 IYEEETDWINIHIDIFHVFPOSHEEIEFIFASECTGFRHLKYTISI 469
 DB 421 IYEEETDWINIHIDIFHVFPOSHEEIEFIFASECTGFRHLKYTISI 469

RESULT 10

Q9HBM3 PRELIMINARY; PRT; 465 AA.
 AC Q9HBM3;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Dipeptidyl peptidase 8 (Fragment).
 GN DPP8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=20467194; PubMed=11012666;
 RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
 RA Gortel M.D.;
 RA "Cloning, expression and chromosomal localization of a novel human
 RT dipeptidyl peptidase (DPP) IV homology, DPP8";
 RL Eur. J. Biochem. 267:6140-6150 (2000).
 DR EMBL; AF221636; AAC29768.1; -
 DR MEROPS; S09.018; -
 FT NON_TER 1
 SO SEQUENCE 465 AA; 53197 MW; 229399EC0A4FE29CB CRC64;

Query Match 51.5%; Score 2422; DB 4; Length 465;
 Best Local Similarity 82.3%; Pred. No. 9.2e-180;
 Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 318 TGTANPKVTFKMSIMIDAGRILIDVIXELIQFELLFEGVEIYARAGWTPGKAYMSI 377
 DB 1 TGTANPKVTFKMSIMIDAGRILIDVIXELIQFELLFEGVEIYARAGWTPGKAYMSI 60
 QY 378 LIDRSQRLQIVLISPLFIPVEDDWERQRLIESVDSVPLIYEETDWINIHIDIF 437
 DB 61 LIDRSQRLQIVLISPLFIPVEDDWERQRLIESVDSVPLIYEETDWINIHIDIF 120
 QY 438 HVFPQSHBEIEFIFASECTGFRHLKYTISILKSKYKRSGLPAPSDCKPIKEIA 497
 DB 121 HVFPQSHBEIEFIFASECTGFRHLKYTISILKSKYKRSGLPAPSDCKPIKEIA 180
 QY 498 ITSGEMEVLGHGNSIQVDEVRRLVYEGTKDQSLFELHLYVSVYVPGSEVTRLDRGYSH 557
 DB 181 ITSGEMEVLGHGNSIQVDEVRRLVYEGTKDQSLFELHLYVSVYVPGSEVTRLDRGYSH 240
 QY 558 SCCISQCHDFISKYSNOKNPHCVSLYKSPEDDPCTKKEFWATLIDSGPLPDYTPP 617
 DB 241 SCCISQCHDFISKYSNOKNPHCVSLYKSPEDDPCTKKEFWATLIDSGPLPDYTPP 300
 QY 618 EIFSESTGTFTLYGMLYKXHDLPKPKYPTVLPYGGPQVQVLYNNRFGKVFRLNTLA 677
 DB 301 EIFSESTGTFTLYGMLYKXHDLPKPKYPTVLPYGGPQVQVLYNNRFGKVFRLNTLA 340
 QY 678 SLGVVVVVIDNRSGCHGKLFEGAFKYMQIEIDDOVEGLQYLASRDPIDLRVGIHG 737
 DB 341 ----- 340
 QY 738 WSYGYLSLMAQMQRSDIFRVAIAGAPVTLMIFPDGYTERVYMGHPDQEGYLLGYAM 797
 DB 341 -----VAIAGAPVTLMIFPDGYTERVYMGHPDQEGYLLGYAM 380
 QY 798 QAEKPSPEPNRLLHGFLENVFAFATSIILSELYVAGKPYDQIYPOEHSIRVPSG 857
 DB 381 QAEKPSPEPNRLLHGFLENVFAFATSIILSELYVAGKPYDQIYPOEHSIRVPSG 440
 QY 858 EHYELHLHLQENLSRIALAKYI 882
 DB 441 EHYELHLHLQENLSRIALAKYI 465

RESULT 11

Q8N3F5 PRELIMINARY; PRT; 628 AA.
 AC Q8N3F5;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN DKFZP762M2413.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Melanoma;
 RA Ansoyge W., Winkner U., Mewes H.W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL834376; CAD39039.1; -
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR000379; Ser-estrs_site.
 DR Pfam; PF00326; Peptidase_S9; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SO SEQUENCE 628 AA; 71368 MW; DC0B6A9440507C3B CRC64;

Query Match 44.0%; Score 2068; DB 4; Length 628;
 Best Local Similarity 60.6%; Pred. No. 5e-152;

Matches 380; Conservative 89; Mismatches 156; Indels 2; Gaps 2;

QY 220 IWISNIVTRERRLTYVNELANMEDARSAVATFVLOEBEDRSGYMWCPKAEFTTSG 279
 Db 1 LWNANIEETGEERLTFCHQGLSNVLDDEPKSAGVATFVLOEBEDRGTGVMWCTTASWSE 60
 QY 280 G-KILRIYENDESEVEIHTVSPMLERADSPRYKGTGANRVTFKSEIMIDAG 338
 Db 61 GKTRIRIYEEVDESEVEIHTVSPALERKTDSPRYKGTGANRVTFKSEIMIDAG 120
 QY 339 RIIDVDEKLIQPFELIFEGVYIARAGTPEKGYMSILDRSQTRLOVILSPLELFP 398
 Db 121 KIVSQEKELVOPESLPEKGYIARAGTPEKGYMSILDRSQTRLOVILSPLELFP 180
 QY 399 VEDVMEBQRLIESVDSVTPPLIYEETDININIHDPHVPQSHER-ELFIPASBCK 457
 Db 181 STENEGRLASARAVRNQVYVVEEVTNWINVHDIFFPPNQERTSSAFAPMAR 240
 QY 458 TGFRLHYKTLISLKSRYKSSGGLPAPSDFKPIKEPIATSGMEVILGRGSIQVDE 517
 Db 241 PASAICTSPPELKSQGYDMSPPSPGDEDFKPIKEPIATSGMEVILGRGSIQVDE 300
 QY 518 VRRLVYFEGTDSPLFHHLYVVSYNPGEVTRLTDRGYSHSCCISOHCDFPISKYNOQN 577
 Db 301 ETKLVYFQGTQDTPLEHHLVVSYEAGEIVALTFRGSHSCMGNPDMFVSHSSVST 360
 QY 578 PHCVSLYKLSPEDDPTCKTEFWATILDSAGPLPDYTPPELIFSESTTGFTLYGMLYKP 637
 Db 361 PCVAVHYKLSGDDDPLEHHLVVSYEAGEIVALTFRGSHSCMGNPDMFVSHSSVST 420
 QY 638 HDLOGKKPTVLTFTYGGPOVOLVNNRFGVRYFFLNTLASIGYVAVVINDNGSCHRGK 697
 Db 421 HALQGGKHPVLTFTYGGPOVOLVNNRFGVRYFFLNTLASIGYVAVVINDNGSCHRGK 480
 QY 698 FEGAKRYMGQIEIDQVEGLQYLASRYPIDLDRYGHWSYGYLSLMALMORSIDIFR 757
 Db 481 FEGALKNMGQVEIDQVEGLQVFAEKYFIDLSVAIHGWSYGYLSLMALMORSIDIFR 540
 QY 758 VAIAGAPVTLMTFTYGTERTYMGHDQNEGGYVIGSVAMQAEKPESENNLLHGFID 817
 Db 541 VAIAGAPVTLMTFTYGTERTYMGHDQNEGGYVIGSVAMQAEKPESENNLLHGFID 600
 QY 818 ENVHFAHTSILSFLVRACKPYDLOIY 844
 Db 601 ENVHFAHTSILSFLVRACKPYDLOIY 627

RESULT 12
 Q9HBM2 PRELIMINARY; PRT: 360 AA.

AC Q9HBM2; 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Dipeptidyl peptidase 8 (Fragment).
 GN DPP8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20467194; PubMed=11012666;
 RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
 RA Gorrell M.D.;
 RT "Cloning, expression and chromosomal localization of a novel human
 RT dipeptidyl peptidase (DPP) IV homolog, DPP8.";
 RL Eur. J. Biochem. 267:6140-6150(2000).
 DR EMBL: AF221637; AAG29769.1; --
 DR MEROPS: S09.018; --
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR000379; Set_estrs_site.
 DR Pfam: PF00326; Peptidase_S9; 1.

FT NON TER 1 1
 FT NON TER 360 360
 FT SEQUENCE 360 AA; 41070 MW; CF81COB61423E22 CRC64;
 Query Match 39.1%; Score 1836.5; DB 4; Length 360;
 Best Local Similarity 67.5%; Pred. No. 2.2e-134; Indels 173; Gaps 1;
 Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

QY 244 EDARSAVATFVLOEBEDRSGYMWCPKAEFTTSGKILRIYENDESEVEIHTVSP 303
 Db 1 EDARSAVATFVLOEBEDRSGYMWCPKAEFTTSGKILRIYENDESEVEIHTVSP 60
 QY 304 MLETRRADSPRYKGTGANRVTFKSEIMIDABGRITDIVDKELQPFELIFEGVEYIA 363
 Db 61 MLETRRADSPRYKGTGANRVTFKSEIMIDABGRITDIVDKELQPFELIFEGVEYIA 98
 QY 364 RAGWTEPKGYMSILDRSQTRLOVILSPLELFIIVEDVMBQRLIESVDSVTPPLIY 423
 Db 99 ----- 98
 QY 424 EETDQWINIHDPHVPQSHERELFIPASECKTGRHLKYTSLKSKYKSSGGLP 483
 Db 99 ----- 98
 QY 484 APSDFKPIKEPIATSGMEVILGRGSIQVDEYRRLVYFEGTDSPLFHHLYVVSYN 543
 Db 99 ----- 127
 QY 544 PGEVTRLTDRGYSHSCCISOHCDFPISKYNOQNPHCVSLYKLSPEDDPTCKTEFWAT 603
 Db 128 PGEVTRLTDRGYSHSCCISOHCDFPISKYNOQNPHCVSLYKLSPEDDPTCKTEFWAT 187
 QY 604 ILDSAGPLPDYTPPELIFSESTTGFTLYGMLYKPHDLOPKKYPTVLTFTYGGPOVOLVNN 663
 Db 188 ILDSAGPLPDYTPPELIFSESTTGFTLYGMLYKPHDLOPKKYPTVLTFTYGGPOVOLVNN 247
 QY 664 RPKGYRYRLNTLASIGYVAVVINDNGSCHRGKKEGAFKRYMGQIEIDQVEGLQYLAS 723
 Db 248 RPKGYRYRLNTLASIGYVAVVINDNGSCHRGKKEGAFKRYMGQIEIDQVEGLQYLAS 307
 QY 724 RYDFDLDRVGHGWSYGYLSLMALMORSIDIFRAVAGAPVTLMTFTYGTGT 776
 Db 308 RYDFDLDRVGHGWSYGYLSLMALMORSIDIFRAVAGAPVTLMTFTYGTGT 360

RESULT 13
 Q9HBM4 PRELIMINARY; PRT: 310 AA.

AC Q9HBM4; 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Dipeptidyl peptidase 8 (Fragment).
 GN DPP8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=placenta;
 RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
 RA Gorrell M.D.;
 RT "Cloning, expression and chromosomal localization of a novel human
 RT dipeptidyl peptidase (DPP) IV homolog, DPP8.";
 RL Eur. J. Biochem. 267:6140-6150(2000).
 DR EMBL: AF221635; AAG29767.1; --
 DR MEROPS: S09.018; --
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR00326; Peptidase_S9; 1.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR NON_TER 1 1
 SQ SEQUENCE 310 AA; 35396 MW; 8B87C34026D9C7AC CRC64;

Query Match 35.0%; Score 1645.5; DB 4; Length 310;
Best Local Similarity 86.4%; Pred. No. 1.2e-119;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY 524 FEETKSPLEHLYVSVYVNGEVTRLDRGVSHSCCISQHCDFPISKYNOKNPKCVSL 583
DB 1 FEGTKSPLEHLYVSVYVNGEVTRLDRGVSHSCCISQHCDFPISKYNOKNPKCVSL 60
QY 584 YKLSPEDEPTCKTEFWATILDSAGPLDYPPEIIFSESTGTFTLYGMLYKPHLOPG 643
DB 61 YKLSPEDEPTCKTEFWATILDSAGPLDYPPEIIFSESTGTFTLYGMLYKPHLOPG 120
QY 644 KKPFTLYFYGGQVQVUNNRKGVKFRNLNTASLGIVVYVINDNRGSHRGKFGAEX 703
DB 121 KKPFTLYFYGGQVQVUNNRKGVKFRNLNTASLGIVVYVINDNRGSHRGKFGAEX 134
QY 704 YKMGQIEIDDOVEGLQYIASRYDITLDRVGIHGSYGGLSLMALMORSIDIFRVAIAGA 763
DB 135 ---GQIEIDDOVEGLQYIASRYDITLDRVGIHGSYGGLSLMALMORSIDIFRVAIAGA 191
QY 764 PTLWTFYDTGYTERYMGHPDQNEQGYVIGSVAMAQAEKFPSEPNRLLLHGFLENVHFA 823
DB 192 PTLWTFYDTGYTERYMGHPDQNEQGYVIGSVAMAQAEKFPSEPNRLLLHGFLENVHFA 251
QY 824 HRSILSLFVPAKRPYDLOIYQERHSIRVPSGGEHYELHLHYLOENTGSRIALKVI 882
DB 252 HRSILSLFVPAKRPYDLOIYQERHSIRVPSGGEHYELHLHYLOENTGSRIALKVI 310

RESULT 14
075273 PRELIMINARY; PRT; 508 AA.

AC 075273;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE R26984.1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan W., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 2.5 Mb region in 19p13.3";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005594; AAC33801.1; -
DR MEROPS; S09.019; -
DR InterPro; IPR000379; Set_eestrs_site.
FT NON_TER 1
SQ SEQUENCE 508 AA; 57750 MM; 2FAD6645BE2D2C89 CRC64;

Query Match 35.0%; Score 1643.5; DB 4; Length 508;
Best Local Similarity 57.6%; Pred. No. 3.8e-119;
Matches 313; Conservative 77; Mismatches 108; Indels 45; Gaps 6;

QY 340 TIDVIDKELIQPFELIFEGVEYIARAGWTPECKVAMSLIDRSQTRLOIVLISPELFTPV 399
DB 1 IVSTDEKELVQPFSSLPKVEYIARAGWTPECKVAMSLIDRSQTRLOIVLISPELFTPV 53
QY 400 EDDVNERQLLESVDSTPLIVBETDININIDITHVPOSH-BEIEIFIFASECKT 458
DB 54 TENEQRIASAPARVPRNVQYVVEVNVINVDIYPPQSGEDELFLRLANCKT 113
QY 459 GFRHLKITSILKSKYKSSGGLPAPSDFKCPIKEIATISGEWVILGRHSNIGVDEV 518

DB 114 GFCHLYKVAVLKSGQYDWSPPSPG-----EGEQSLTNAIW-----VNEE 154
QY 519 RRLVYEGTKDPSLEHLYVSVYVNGEVTRLDRGVSHSCCISQHCDFPISKYNOKN 578
DB 155 TKLVYFGTKDPSLEHLYVSVYVNGEVTRLDRGVSHSCCISQHCDFPISKYNOKN 214
QY 579 HCVSLKYSPEDEPTCKTEFWATILDSAGPLDYPPEIIFSESTGTFTLYGMLYKPH 638
DB 215 PCHVYKLSGPDDEPLHKGPRFWSAMEAA-----KIFPHRSVDRLIGMITYKP 265
QY 639 DLQPGKRYTFLFYGGQVQVUNNRKGVKFRNLNTASLGIVVYVINDNRGSHRGK 698
DB 266 ALQPGKRYTFLFYGGQVQVUNNRKGVKFRNLNTASLGIVVYVINDNRGSHRGK 325
QY 699 EGAFKYMGOIEIDDOVEGLQYIASRYDITLDRVGIHGSYGGLSLMALMORSIDIFR 758
DB 326 EGALKNMQQVELEDDVEGLQYIAEYGEFIDLSRAVHGSYGGLSLMALMORSIDIFR 385
QY 759 ALAGAPVTLMIFYDTGYTERYMGHPDQNEQGYVIGSVAMAQAEKFPSEPNRLLLHGF 818
DB 386 ALAGAPVTLMIFYDTGYTERYMGHPDQNEQGYVIGSVAMAQAEKFPSEPNRLLLHGF 445
QY 819 NVHFAHTSILSLFVPAKRPYDLOIYQERHSIRVPSGGEHYELHLHYLOENTGS 869
DB 446 NVHFAHTSILSLFVPAKRPYDLOIYQERHSIRVPSGGEHYELHLHYLOENTGS 505
QY 870 ENL 872
DB 506 EYL 508

RESULT 15
09BVR3 PRELIMINARY; PRT; 439 AA.

AC 09BVR3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strusberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC000970; AAH00970.1; -
DR MEROPS; S09.019; -
DR GeneW; HGNC:18648; DPP9.
DR InterPro; IPR001375; Peptidase_S9.
DR DR InterPro; IPR000379; Set_eestrs_site.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 439 AA; 49926 MM; A18BBA9E12092BAF CRC64;

Query Match 34.3%; Score 1610.5; DB 4; Length 439;
Best Local Similarity 67.0%; Pred. No. 1.1e-116;
Matches 294; Conservative 57; Mismatches 87; Indels 1; Gaps 1;

QY 435 DIFHVFQPSH-BEIEIFIFASECKTGFRLHYKITSILKSKYKSSGGLPAPSDFKCP 493
DB 1 DIFHVFQPSH-BEIEIFIFASECKTGFRLHYKITSILKSKYKSSGGLPAPSDFKCP 60
QY 494 BEIATISGEWVILGRHSNIGVDEVRLVYEGTKDPSLEHLYVSVYVNGEVTRLD 553
DB 61 BEIATISGEWVILGRHSNIGVDEVRLVYEGTKDPSLEHLYVSVYVNGEVTRLD 120
QY 554 GYSHSCCISQHCDFPISKYNOKNPKCVSLYKLSPEDEPTCKTEFWATILDSAGPLD 613

```

Db      121 GFSHSCSMQNFDMFVSHYSSVSTPPCVHYKLSGPDDDLHKQPRFPMASMEASCPD 180
QY      614 YFPPEIFSEBESTGFTLYGMLYKPHDLQPKKXPTVLFYGGPOVLVNNRFKGYEFL 673
Db      181 YFPPEIFHFHTRSDVRLYGMLYKPHALOPEKKHPTLVFYGGPOVLVNNSPFKGIXYRL 240
QY      674 NTLASIGYVVVVIDNRSCHRGKFEAGFYKMGQLEIDDOVEGLQYLAARYDFIDDRY 733
Db      241 NTLASIGYAVVVIDGRSCQGRGLRFBGALKNQMGQVEIEDQVEGLQFAEKYGFIDLSRY 300
QY      734 GTHGMSYGYLSLMLMQRSDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLG 793
Db      301 AIHMSYGGFLSLMGLIHKPQVFKVAIAGAPVTVMAYDTGYTERYMDVPENNQHGYEAG 360
QY      794 SVAMQABKPPSEPNRLLILHGFIDENYHFAHTSILSFLVRAGKPYDLOIYQERHSIRY 853
Db      361 SVLHYVEKLEPNERNRLLILHGFIDENYHFFHNFVLSQILIRAGKPYOLOIYPNERSHIRC 420
QY      854 PESGEHYEHLHLHYLOENTL 872
Db      421 PESGEHYEVTLLHFLQBYL 439

```

Search completed: October 15, 2003, 17:13:41
 Job time : 61.4095 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: October 15, 2003, 17:15:46 ; Search time 9629.41 seconds
(without alignments)

3747.092 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700
Sequence: 1 MAAMETBQGVHIFETADC.....HLHYLOENIGSRITALKVI 882

Scoring table:
BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Rgapop 6.0 , Rgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=x1h
-Q=/cgm2.1/USPRO.spool/US10070464/runat.15102003.113553.24810/app.query.fasta.1.2652
-DB=Genembl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCR=0 -LOOPEXT=0
-UNITS=bits -START=1 END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464@cgm2.1.13561@runat.15102003.113553.24810 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -XLANGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Genembl:*
1: gb Da:*
2: gb Hcg:*
3: gb In:*
4: gb Om:*
5: gb Ov:*
6: gb Pat:*
7: gb Ph:*
8: gb Pl:*
9: gb Pr:*
10: gb Ro:*
11: gb Sts:*
12: gb Sy:*
13: gb Un:*
14: gb Vi:*
15: em Ba:*
16: em Fun:*
17: em Hum:*
18: em In:*
19: em Mu:*
20: em Om:*
21: em Or:*
22: em Ov:*
23: em Pac:*
24: em Ph:*
25: em Pl:*
26: em Ro:*
27: em Sts:*
28: em Un:*

29: em Vi:*
30: em Htg Hum:*
31: em Htg Inv:*
32: em Htg Other:*
33: em Htg Mus:*
34: em Htg Pln:*
35: em Htg Rod:*
36: em Htg Mam:*
37: em Htg Vrt:*
38: em Gy:*
39: em Htgo Hum:*
40: em Htgo Mus:*
41: em Htgo Other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4700	100.0	2649	6 AX354795	AX354795 Sequence
2	4700	100.0	2656	9 AY172659	AY172659 Homo sapi
3	4700	100.0	2671	6 AX608725	AX608725 Sequence
4	4700	100.0	3106	6 AX342633	AX342633 Sequence
5	4700	100.0	3127	9 AF221634	AF221634 Homo sapi
6	4700	100.0	3143	6 AX354793	AX354793 Sequence
7	4696	99.9	4535	9 BC040203	BC040203 Homo sapi
8	4680	99.6	4829	6 AX608735	AX608735 Sequence
9	4536.5	96.5	4799	10 BC043124	BC043124 Mus muscu
10	4397.5	93.6	3130	6 BC030688	BC030688 Homo sapi
11	4385.5	93.3	4685	6 AX608745	AX608745 Sequence
12	4385	93.3	4676	6 AX608743	AX608743 Sequence
13	4118	87.6	2842	6 AX405770	AX405770 Sequence
14	4092.5	87.1	4523	6 AX608731	AX608731 Sequence
15	4091	87.0	2778	6 AK000290	AK000290 Homo sapi
16	3970.5	84.5	2510	6 AX338497	AX338497 Sequence
17	3771	80.2	2668	6 AX405771	AX405771 Sequence
18	3661.5	77.9	4309	6 AX608737	AX608737 Sequence
19	3364.5	71.6	2161	6 BD157001	BD157001 Primer fo
20	3364.5	71.6	2161	9 AK027826	AK027826 Homo sapi
21	2870	61.1	2602	9 AY172660	AY172660 Homo sapi
22	2870	61.1	2617	6 AX608727	AX608727 Sequence
23	2870	61.1	3716	6 AX480934	AX480934 Sequence
24	2870	61.1	4219	6 AX608751	AX608751 Sequence
25	2870	61.1	4232	9 BC037948	BC037948 Homo sapi
26	2870	61.1	4295	9 AF452102	AF452102 Homo sapi
27	2870	61.1	4302	6 AX608747	AX608747 Sequence
28	2867	61.0	2676	6 AX524928	AX524928 Sequence
29	2863	60.9	3000	9 AF542510	AF542510 Homo sapi
30	2820.5	60.0	4180	6 AX608759	AX608759 Sequence
31	2820.5	60.0	4263	6 AX608757	AX608757 Sequence
32	2649	56.4	4076	6 AX608755	AX608755 Sequence
33	2649	56.4	4159	6 AX608753	AX608753 Sequence
34	2599.5	55.3	4037	6 AX608763	AX608763 Sequence
35	2599.5	55.3	4120	6 AX608761	AX608761 Sequence
36	2422	51.5	1669	9 AF221636	AF221636 Homo sapi
37	2139.5	45.5	3243	9 HSM80548	ALH834576 Homo sapi
38	1836.5	39.1	1083	9 AF221637	AF221637 Homo sapi
39	1730.5	36.8	2261	6 AX713363	AX713363 Sequence
40	1730.5	36.8	2261	9 AK054656	AK054656 Homo sapi
41	1645.5	35.0	1197	9 AF221635	AF221635 Homo sapi
42	1610.5	34.3	2546	9 BC000970	BC000970 Homo sapi
43	1596	34.0	3740	3 BT001499	BT001499 Drosophi1
44	1391	29.6	2411	6 AX608749	AX608749 Sequence
45	1360	28.9	215711	3 AC008205	AC008205 Drosophi1

RESULT 1

ALIGNMENTS

AX354795
LOCUS AX354795 2649 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 3 from Patent WO0179473.
ACCESSION AX354795
VERSION AX354795.1 GI:18619528
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Meyers, R.A. and Williamson, M.
21953, a human prolyl oligopeptidase family member and uses thereof
Patent: WO 0179473-A 3 25-OCT-2001;
JOURNAL Millennium Pharmaceuticals, Inc. (US)
FEATURES
source 1..2649
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 803 a 514 c 586 g 746 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2649
Score: 4700.00 Matches: 882
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-070-464-1 (1-882) x AX354795 (1-2649)
QY 1 MetAlaAlaAlaMetGluThrGluGluLeuGluValGluIlePheGluThrAlaAspCys 20
DB 1 ATGGAGAGAGATGAGAAACAGAACAGCTGGGTGTGATGATTTGAACTCCGACCTGT 60
QY 21 GluGluAsnIleGluSerGlnAspArgProIleuGluProPheTyrValGluArgTyr 40
DB 61 GAGGAGATATTGATGATCAAGATCGGCTTAATTGAGCCCTTTTATGTTGAGGCGGTAT 120
QY 41 SerTPSerGlnLeuValLeuLeuAlaAspThrArgGlySerHisGlyTyrMetMet 60
DB 121 TCCTGGAGTCACTTAAAGCTGCTTCCCATACCAAAATATCATGCTACATGATG 180
QY 61 AlaAlaAlaProHisAspPheMetPheValIleSarAsnAspProAspGlyProHisSer 80
DB 181 GCTAAGGACACCATGATTCATGTTGTGAAGAGATGATCCAGATGACCTCATTTCA 240
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
DB 241 GACGAAATCTATTAACCTTGCATGCTGTGTGAAGAGATGATCCAGATGACCTCATTTCT 300
QY 101 GluIleProIleThrIleAsnArgAlaAlaValLeuMetLeuSerTyrPheProLeuLeu 120
DB 301 GAATTTCCCAAAACATATCAATAGACAGACATCTTAAGCTCTCTTGAAGCCCTTTTG 360
QY 121 AspleuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
DB 361 GATCTTTTTCAGGACACCTGACCTATGGAATGTAATCTCGAAGAAAGAACTTTAAGA 420
QY 141 GluArgIleAspArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
DB 421 GAAAGAAACCGCATTTGAAACAGCTGGAAATTCCTTCTTCATTCACCAAGAGAGTGA 480
QY 161 ThrPheLeuPheGlnAlaGlySerGlyTyrHisValIleAspGlyGlyProGlnGly 180
DB 481 ACATTTCTGTTTCAAGCCGATGAGATTTATCATCGTAAAGATGAGGCGCACAAAGA 540
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 541 TTAAAGCAACACTTTAAGGCCCAATCTAGTGGAACACTAGTTGTCCCAACATACGATG 600

QY 201 AspProIleuGluCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
DB 601 GATCAAAATATATGCGCTGATCCAGACCTGATGCTTTTATACATAGCAACATATT 660
QY 221 TyrIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
DB 661 TGGATATCTAACATCTGTAAACAGAGAAAGAGAACTCATTAATGTCACAAATGAGCTA 720
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluVal 260
DB 721 GCCAACATGAGAGAAAGATGCCAGATCAGCTGAGTCCCTGCTTGTCTCCAAAGAGAA 780
QY 261 PheAspArgTyrSerGlyTyrTyrTyrProCysProIleAlaGluThrThrProSerGlyGly 280
DB 781 TTGATATGATATTTCTGCTGATTTGATGATGATCCAAAGCTGAAACACTCCAGTGTGT 840
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
DB 841 AAAATTTCTGAAATTCATATATGAAGAAATGATGATCTGAGTGAATTAATTCATGTT 900
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProIleThrGlyThr 320
DB 901 ACATCCCTATGTTGAGAAACAGAGAGGCAATTCATTCCTGTTATCTTAAACAGATACA 960
QY 321 AlaAsnProIleValThrPheLeuMetSerGluIleMetIleAspAlaGluGlyArgIle 340
DB 961 GCAATCTCTAAAGCACCTTTAAGATGTCAGAAATTAATGATGATGTCGAAAGAGATC 1020
QY 341 IleAspValIleAspIleGluLeuIleGluProPheGluIleLeuPheGluGlyValGlu 360
DB 1021 ATAGATGTCATGATGAAGAGAACTAATTCACCTTTTGATTCATATTGAAAGAGTTGAA 1080
QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyIleValTyrAlaTyrSerIleLeuLeuAsp 380
DB 1081 TATATTGCCAGAGCTGATGATGATCTCTGAGGAGAAATATGCTTGTGTCATCTTACAT 1140
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
DB 1141 CGTCCCAAGACTCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
QY 401 AspAspValMetGluGluGluGluArgLeuIleGluSerValProAspSerValThrProLeu 420
DB 1201 GATGATGTTATGAGAAAGCAGAGACATTAATGATGATGATGATGATGATGATGATGAT 1260
QY 421 IleIleTyrGluGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
DB 1261 ATTATCTATGAGAAACAGACAGATCTGATTAATATCCATGATATTTTCACTTTT 1320
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIleThrGlyPhe 460
DB 1321 CCCCAAGATCAGAAAGAGAAATGAGTTATTTTGGCTCTGAATGCAAAACAGTTTC 1380
QY 461 ArgHisLeuTyrIleThrSerIleLeuIleGluSerIleTyrIleArgSerSerGly 480
DB 1381 CGTATTTATACAAATTAATCTATTTTAAAGAGAAATTAATTAATTAATTAATTAATTA 1440
QY 481 GlyLeuProAlaProSerAspPheLeuCysProIleLeuGluGluIleAlaIleThrSer 500
DB 1441 GGCTGCTGCTGCTCAAGATGATTAATGATGATGATGATGATGATGATGATGATGATG 1500
QY 501 GlyIleTyrProIleValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
DB 1501 GGTAAATGAGAGAGTTCTTGGCCGATGATCTAATATCCAAAGTTGATGAAGTACAGAG 1560
QY 521 LeuValTyrPheGluGluGlyThrIleAspSerProLeuGlnHisIleLeuTyrValIleSer 540
DB 1561 CTGCTATATTTTGAAGCACCAAAAGATCCCTTTAAGCATCACTGTACGTGTCAGT 1620
QY 541 TyrValAsnProGluGluValThrArgLeuThrAspArgIleTyrSerHisSerCysCys 560
DB 1621 TACGTAATCTCTGAGAGGTGACAAAGGCTGACCTGCTGCTGCTGCTGCTGCTGCTG 1680
QY 561 IleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnIleAsnProHisCys 580

```

Db      1681 ATCAGTACGACCTGACTCTTTATATAGTAGTATGTAACCAAGAAATCCACACTGT 1740
Qy      581 ValSerLeuTyrIleuSerSerProGluAspAspProThrCysIleuThrIleuGluPhe 600
Db      1741 GTGTCCTTTACAGAGCTATCAAGCTCGTAAGATGACCCAACTTGCAAAACAAGAAATTT 1800
Qy      601 TTPAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db      1801 TGGGCCACCAATTTGGATTTCAGCAGAGTCTCTCTGACTATCTCTCCCAAGAAATTTTC 1860
Qy      621 SerPheGluSerThrThrGluPheThrIleuTyrGlyMetLeuTyrIleuProIleAspLeu 640
Db      1861 TCTTTTGAAGACTACTGAGATTGATTCATTGATGGAGATGCTTCAAGAGCTTCATGATCTTA 1920
Qy      641 GluProGlyIleuTyrIleuProThrValLeuPheIleTyrGlyGlyProGluIleValGluLeu 660
Db      1921 CAGCTCGAAAGAAATATCTACTGCTGCTGTTCAATATATGTTGTTCTCCAGGTGCAAGTTG 1980
Qy      661 ValAsnAsnArgPheIleuGlyValIleuTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      1981 GTGATATATCGGTTTAAAGAGTCAGATATTTCCGCTTGATATACCTTAGCTCTCTAGGT 2040
Qy      681 TyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuIleuPheGluGly 700
Db      2041 TATGTGTTGATGATGATGACACAGAGGATCTCTGTACCGAGGCTTAAATTTGAAGGC 2100
Qy      701 AlaPheIleTyrIleuMetGlyGluIleGluIleAspAspGluValGluIleuGluIle 720
Db      2101 GCCTTTAAATATATAATGCGGTCAATATGAAATTCAGCATCGGTGGAAGCATCCCAATAT 2160
Qy      721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyr 740
Db      2161 CTAGCTTCTCGATATGATTTTCATTCATGACTTATGATGTGTGGGATCCACGCGTGTCTCTAT 2220
Qy      741 GlyIleTyrIleuSerIleuMetAlaLeuMetGluIleArgSerAspIlePheArgValAlaIle 760
Db      2221 GGAGGATACCTCTCTCGATGAGCTTATATGACAGAGCTCAGATATCTTCAGGGTTGCTAAT 2280
Qy      761 AlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db      2281 GCTGGGGCCCCAGTCACTCTGTGATCTTCTATGATACAGATACAGGAACGTTATATG 2340
Qy      781 GlyHisProAspGluAsnGluGluIleTyrTyrIleuGlySerValAlaMetGluAlaGlu 800
Db      2341 GGTCACCCCTGACAGATGAAACAGGGCTATTATCTAGGATCTGTGGCCATGCAAGCAGAA 2400
Qy      801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db      2401 AAGTTCCTCTGACCAAAATCGTTTACTGCTCTTAATGTTTCTCGATGAGATGTC 2460
Qy      821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaArgAlaGlyLysProTyrAsp 840
Db      2461 CATTTTGCATATCCAGTATATATTAATGATTTTATGAGGGCTGGAAGCCATATGAT 2520
Qy      841 LeuGluIleTyrProGluIleuArgHisSerIleArgValProGluSerGlyGluHisTyr 860
Db      2521 TTACAGATCTATCTCAGAGAGACACAGATTAAGATTCCTGATGAGGAACATTAAT 2580
Qy      861 GluLeuHisIleuLeuHisTyrIleuGluGluIleAsnLeuGlySerArgIleAlaIleLeuLys 880
Db      2581 GAACCTGATCTTTTGGCACTACCTTCAAGAAACCTTGATCAAGTATTCCTGCTCTTAAA 2640
Qy      881 ValIle 882
Db      2641 GTGATA 2646

```

```

RESULT 2
LOCUS   AY172659      2656 bp      mRNA      linear      PRI 08-JAN-2003
DEFINITION Homo sapiens dipeptidyl peptidase IV-related protein-1 (DPP1)
mRNA, complete cds.
ACCESSION AY172659

```

```

VERSION AY172659.1 GI:27549549
KEYWORDS
SOURCE  Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2656)
AUTHORS Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
TITLE Novel Serine Protease Genes Related to DPP1v
JOURNAL Patent: US (WO 0231134)-A 18-APR-2002;
2 (bases 1 to 2656)
AUTHORS Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2002) Ferring Research Institute, 3550 General
Atomic Ct., San Diego, CA 92121, USA
FEATURES
source
1..2656
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q22.1-q22.2"
1..2656
/gene="DPP1"
/gene="DPP1"
8..2656
/gene="DPP1"
/codon_start=1
/product="dipeptidyl peptidase IV-related protein-1"
/protein_id="AA017261.1"
/db_xref="GI:27549550"
/translation="MAAETBQLGVEIFETADCEENIESODRKLSEPFVERYSWQ
LKKLADTRRYGYMMAKAPHDPMFVRNPDGSDRIYILAMSGENRENTLFSYSEI
PKTINRAVALMLSMKPLDLDFQALDFPQALREBELRERKIGTVGASVDYQSGG
TFLFOAGSGIYHYVGDGPOGFTOOPRLPNLVETSCPINRMDPKCPADPMWIAFHSN
DWISNIVTREERRLTYVHNEILANMEDASAGATATLOEPRRCYGGYCCPAEAT
PSGKILRIIYENDESEVATHTSMTLTRADSRYPRTGTANPKVTRKSEIMI
DAGRIIIVDIDKLIQPELIFGVEYIAAGWTPREKYAWSILLDSQRLQVLVS
PELFIPEVDVMERORLIESVPDSVPLIYEETTDIMINHDFHVFQSHBEIER
IFASECTGFRHLKYIKSLKESKYKSSGGLPAPSDPKCIKEEIAITSEMEVLGR
HGSNIQVDEVRRLVYEGTKDSPLEHRLVYVSYVNPGEVTRLDRGYSHSCISQHC
FPIKSYNQNCPHCVSLYKLSPEDDPTCKTKEMATILDSAGLPIPTTPETISFES
TTGFTLGMILYKPHDLQPGKRYPLVLTYGQPOVLNNRFGVYFRLLNLAGLY
VVVIDNRGSHRGKEGAPKRYKMGQIIBIDDVQGLQYLASRYPDIDLDRGHSWY
GYLSLMAIMQSRDIFERVALAGAPVTMLIYDTGYTRMYGMHPDNRGVLGSAVAG
AEKFPSEPNRLILHGLFDENHFAHFIISLFLVRAGKPYDLQIYQERSIRVPES
GEHYELHLHYLDENGSRITAAIKVI"
BASE COUNT 804 a 517 c 588 g 747 t
ORIGIN
Alignment Scores:
Pred. No.: 0
Score: 4700.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 9 Gaps: 0
US-10-070-464-1 (1-882) x AY172659 (1-2656)
Qy      1 MetAlaAlaAlaMetGluThrGluGluIleuGluIlePheGluThrAlaAspCys 20
Db      8 ATGCGACGACCAATGGAACAGACAGCTGGGTGAGATATTTGAAGCTGCGGACTGT 67
Qy      21 GluGluAsnIleGluSerGluAspArgProLysLeuGluIleuProPheTyrValIleuArgTyr 40
Db      68 GAGAGAAATTTTAATACACGATCGGCTTAAATTTGAGGCTTTTATGTTGAGCGGATAT 127
Qy      41 SerTyrSerGluIleuLysIleuLeuAlaAspThrArgIleTyrThrIleGlyTyrMetMet 60
Db      128 TCCTGAGAGTACGCTTAAAAAGCTGCTGCGGATACCAAGAAATATCATGCTACATGATG 187
Qy      61 AlaLysAlaProHisAspPheMetPheValIleArgAsnAspProAspArgIlyProHisSer 80

```

Db 188 GCTAAGCAACCATGATTTCATGTTTGTGAAGAGAAATGATCCAGATGACCTCATTTCA 247
Qy AsparGileTYrTYrLeuAlaMetSerGlyGluMetArgGluAsnThrLeuPheTYrSer 100
Db 248 GACAAATCTATTACTTCCATGCTGCTGTGAGAACAGAGAAATACACTGTTTATTTCT 307
Qy 101 GjuilleProLYsThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLYsProLeuLeu 120
Db 308 GAAATTTCCCAAAATCATCATATGACAGACAGCTTAAATGCTCTTGGAAAGCTCTTTTG 367
Qy 121 AspleuPheGlnAlaThrLeuAspTYrGlyMetTYrSerArgGluGluLeuLeuArg 140
Db 368 GATCTTTTTCAGGCAACACTGAGCTATGAGATGATTTCTCGAAGAAAGAACTATTAAAGA 427
Qy 141 GjuArgLYsArgIleGlyThrValGlyIleAlaSerTYrAspTYrHisGlnLYsGly 160
Db 428 GAAAGAAAACGCACTTGGAAACACTCCGAATTCCTTACGATTAACACCAAGAAAGTGA 487
Qy 161 ThrPheLeuPheGlnAlaGlySerGlyIleTYrHisValLYsAspGlyGlyProGlnGly 180
Db 488 ACATTTCTGTTCAGACCGGTAGTGAAATTTATCATGTAAGATGAGAGGCCACAAAGGA 547
Qy 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db 548 TTTACGCAACAACTTTAAGGCCATCTAGTGAACCTAGTTGCCAACATACAGAGATG 607
Qy 201 AspProLYsLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
Db 608 GATCCAAATTTATGCCCTGCTGATCCAGACTGGAATTGCTTTATACATAGCAACAGATTT 667
Qy 221 TrpIleSerAsnIleValThrArgGluGluValArgGlyLeuThrTYrValHisGlnGluLeu 240
Db 668 TGGATATCTAATCATGTAACCAAGAAAGAAAGAAACCTCATATGTCACATAGACTA 727
Qy 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db 728 GCCAACATCGAAGAAAGATGCCAGATCACTGAGTGCCTTCTTGTCTCCAAAGAAAGA 787
Qy 261 PheAspArgLYsSerGlyTYrTrpTrpCysProLYsAlaGluThrThrProSerGlyGly 280
Db 788 TTTGTATGATATTTCGGCTATTGGGTGTTGTCGTAAGAAAGCTGAACAACTCCCAAGTGT 847
Qy 281 LysIleLeuArgIleLeuTYrGluGluAsnAspGlySerGlyValGluIleHisVal 300
Db 848 AAAATTTCTTAAGATTTCTATGTAAGAAATGATGATGATGAGTGAATTAATTCATCTT 907
Qy 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTYrProLYsThrGlyThr 320
Db 908 ACATCCCTATCTTGAGAAACAAAGAGGCAATTCATTCCTTATCTTAAACAGATACA 967
Qy 321 AlaAsnProLYsValThrPheLYsMetSerGlyIleMetIleAspAlaGluGlyArgIle 340
Db 968 GCAATATCTTAAGCTACTTTAAGATGTCAGAAATTAATGATGATGCTGAAGAGAAAGTC 1027
Qy 341 IleAspValIleAspLYsGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db 1028 ATAGATGTCATAGATAGAACTAATTAACCTTTTGAGATCTAATTTGAAGAGGTTGAA 1087
Qy 361 TYrIleAlaArgAlaGlyTYrThrProGluGlyLYsTYrAlaTrpSerIleLeuLeuAsp 380
Db 1088 TATATTTGCCAGAGCTGGATGAGCTCCTGAGGGAATAATGCTTGGTCCATCTTACTAGAT 1147
Qy 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db 1148 CGCTCCCAAGACTCCCTGCAATAGTGTGATCTCACCTGAAATTAATTAATCCAGTGA 1207
Qy 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db 1208 GATATGTTATGAGAAAGCAGAGACTCATTAAGTCAAGCTGATTTCTGTAAGCCACTA 1267
Qy 421 IleIleTYrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
Db 1268 ATTATCTATGAGAAACAAAGACATCTGATTAATATCATGACATCTTTCAATGTTTTT 1327

Qy 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeuThrGlyPhe 460
Db 1328 CCCAAATCTACAGAAAGAGAAATTCAGTTTATTTTTCCTCTGATGCAAAACAGCTTTC 1387
Qy 461 ArgHisLeuTYrLYsIleThrSerIleLeuLYsGluSerLYsTYrLYsArgSerGly 480
Db 1388 GCTCATTTATACAAATATACATCTATTATTAAGGAAACAAATATTAACGATCCAGTGT 1447
Qy 481 GlyLeuProAlaProSerAspPheLYsCysProIleLYsGluGluIleAlaIleThrSer 500
Db 1448 GGGCTGCTGCTCCAAAGATTTCAAGTGTCTTATCAAAAGAGAGATGCAATTCAGAT 1507
Qy 501 GlyGluTrpGluValLeuGlnArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db 1508 GGTGAATGGAGATTTCTTGCCCGCATGTATTAATATCCAAAGTTGATGAGTCAAGAGG 1567
Qy 521 LeuValTYrPheGluGlyTYrLYsAspSerProLeuGlnHisIleLeuTYrValIleSer 540
Db 1568 CTGGTATATTTTGAAGGCAACCAAGACTCCCTTTAGAGCATCACTGTAGCTAGTCACT 1627
Qy 541 TYrValAsnProGlyGluValThrArgLeuThrAspArgLYrTYrSerHisSerCys 560
Db 1628 TACGTAATCTCTGAGAGGTGACAAAGCTGACTGACCTGTGCTACTCATCTTCTGTGC 1687
Qy 561 IleSerGlnHisCysAspPhePheIleSerLYsTYrSerAsnGlnLYsAspProHisCys 580
Db 1688 ATCAGTCAAGCACTGATCTTTTATAGTAAGTATGATTAACCAAGAAATCCACACTGT 1747
Qy 581 ValSerLeuTYrLYsLeuSerSerProGluAspAspProThrCysLYsThrLYsGluPhe 600
Db 1748 GTGTCCCTTTCAAGCTATCAAGTCTGAAAGATGACCCAACTTGCAAAACAAAGGAATTT 1807
Qy 601 TrpIleThrIleLeuAspSerAlaGlyProLeuProAspTYrThrProProGluIlePhe 620
Db 1808 TGGGCTACCATTTTGGATTCAGCAGGTCTCTTCTGACTATACCTCCCAAGAAATTTTC 1867
Qy 621 SerPheGluSerThrThrGlyPheThrLeuTYrGlyMetLeuTYrLYsProHisAspLeu 640
Db 1868 TCTTTTGAATATCACTACGATTTACATGATGAGATGCTCTACCAAGCTCATGATCTTA 1927
Qy 641 GlnProGlyLYsLYsTYrProThrValLeuPheIleTYrGlyGlyProGlnValGlnLeu 660
Db 1928 CAGCTCGAAAGAAATATCTCACTGCTGCTGCTCATATGATGTTGCTCAGGTGAGTTG 1987
Qy 661 ValAsnAsnArgPheLYsGlyValLYsTYrPheArgGluAsnThrLeuAlaSerLeuGly 680
Db 1988 GTGAATATCATTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTCTAGGT 2047
Qy 681 TYrValValValIleAspAsnArgLYsSerCysHisArgGlyLeuLYsPheGluGly 700
Db 2048 TATGTGTTGATGATGATGACAAACAGGGATCTGTCAACCAAGGCTTAAATTTGAAAGC 2107
Qy 701 AlaPheLYsTYrLYsMetGlyGlnIleGlnIleAspAspGlnValGlyLYsLeuGlnTYr 720
Db 2108 GCCTTTAAATTTAAAGAGGTCAATAGAAATTAACATCAGCTGAGAAAGCACTCAATAT 2167
Qy 721 LeuAlaSerArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTYr 740
Db 2168 CTACTCTTCGATATGATTTTTCATTTGATGATGATGATGATGATGATGATGATGATGAT 2227
Qy 741 GlyLYsTYrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db 2228 GAGAGATACCTCTCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2287
Qy 761 AlaGlyAlaProValThrLeuTrpIlePheTYrAspThrGlyTYrThrGluArgTYrMet 780
Db 2288 GCTGGGGCCCAAGCACTGTGTGATCTTATATATCAGAGTACAGAGAACCTTATATG 2347
Qy 781 GlyHisProAspGlnArgGluGlnGlyTYrTYrLeuGlySerValAlaMetGlnAlaGlu 800
Db 2348 GGTACCTTCACAGATGACAGGCTATTAATTAAGATCTGTGGCCATGCAAGCAGAA 2407

QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHISGlyPheLeuAspGluAsnVal 820
|
|
Db 2408 AAGTTCCTCTGTGAACCAATCGTTTACTGCTTACATGGTTTCTCGAGTGAACATGTC 2467
|
QY 821 HisPheLeuHisThrSerLeuLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
|
|
Db 2468 CATTTCGACATACACATATATTTACTAGTTTTTTATGAGGGCTGGAACCCATATGAT 2527
|
QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
|
|
Db 2528 TTACAGATCTATCTCTCAGAGAGACACAGCATTAAGAGTCTGTAATCGGAGAACATAT 2587
|
QY 861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaLeuLys 880
|
|
Db 2588 GAATCGATCTTTTGTGACTACCTCAAGAAACCTTGATCAGTATGCTGCTCTAAAA 2647
|
QY 881 ValIle 882
|
|
Db 2648 GTGATA 2653
|
RESULT 3
AX608725 2671 bp DNA linear PAT 17-FEB-2003
LOCUS AX608725
DEFINITION Sequence 2 from Patent WO0231134.
ACCESSION AX608725
VERSION AX608725.1 GI:28404298
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
TITLE Novel serine protease genes related to dppiv
JOURNAL Patent: WO 0231134-A 2 18-APR-2002;
Ferring BV (NL)
FEATURES
source Location/Qualifiers
1..2671
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 805 a 524 c 594 g 748 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2671
Score: 4700.00 Matches: 882
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-070-464-1 (1-882) x AX608725 (1-2671)

QY 101 GluIleProLysThrIleAsnArgAlaIleValLeuMetLeuSerTyrPheProLeuLeu 120
|
|
Db 308 GAAATTCCTCAAACTATCATATAGAGCAGAGCTTTATAGCTCTCTGGAAGCCCTTTTG 367
|
QY 121 AsnLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
|
|
Db 368 GATCTTTTTCAGGACACACAGACTATGATGATTAATCTCGAAGAAAGAACTTTTAA 427
|
QY 141 GluArgHisArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
|
|
Db 428 GAAGAAAGGCAATTGGAACAGTGGAAATTCCTTTACATATTCACCAAGAAAGTGA 487
|
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
|
|
Db 488 ACATTTCTGTTTCAAGCCGGTAGTGAATTTATCACGTTAAAGATGAGGGCCCAAGGA 547
|
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
|
|
Db 548 TTTCAGCAACACCTTTAAGCCCAATCTAGTGAAACTAGTTTCCCAACATACGATG 607
|
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
|
|
Db 608 GATCCAAATTTATGCTCTGCTGATCCAGACTGGAATGCTTTATACATGACAGATATT 667
|
QY 221 TrpIleSerAsnIleValThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240
|
|
Db 668 TGGATATCTAATCATCTGACACAGAGAAAGAGACTGATTTGTCACAAATGACCTA 727
|
QY 241 AlaAsnMetGluGluAsnAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
|
|
Db 728 GCCAACATGGAAGAAATGCGACATGACGTGAGATCCCTACCTTTGTTCCCAAGAA 787
|
QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGly 280
|
|
Db 788 TTTGATGATATTTGCTGCTATTGGTGGTGTCCAAAAGCTGAAACACTCCAGTGGTGT 847
|
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleHisVal 300
|
|
Db 848 AAATTTCTTAAGATTTCTATATGAAGAAATGATGATGATGAGTGAATTTATTCATGTT 907
|
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
|
|
Db 908 ACATCCCTTATGTTGGAACAAAGAGGGCGAGATTCCTGTTATCTTAAACAGTACA 967
|
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
|
|
Db 968 GCAATATCTTAAGTCACTTTTAAAGATGTCAGAAATTAATGATGATGCTGAAGAAAGATC 1027
|
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
|
|
Db 1028 ATGATGTCATATGATTAAGAACTAATCAACCTTTTGAGATTCTATTGAAGAGTTGAA 1087
|
QY 361 TyrIleLeuArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
|
|
Db 1088 TATATTTCCAGAGCTGATGATGATCCCGAAGGAAATATGCTTGCTCATCTCTAGAT 1147
|
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
|
|
Db 1148 CGCTCCAGAGCTGCTCGAGATGATGTTGATTCACCTGAATTTATTTATCCAGTAA 1207
|
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
|
|
Db 1208 GATGATGTTATGGAAGGACAGAGCTCATGAGTCACTGCTGATTCGTGACCCCACTA 1267
|
QY 421 IleIleTyrGluGluThrThrAspIleThrIleAsnIleHisAspIlePheHisValPhe 440
|
|
Db 1268 ATTATCTATGAAGAAACAAAGACATCTGATTAATATTCATGACATCTTTCATGTTT 1327
|
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
|
|
Db 1328 CCCCAAGTCAAGAGAGGAATTTGATTTATTTTCCCTCTGAATGCCAAACAGGTTTC 1387
|

QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
 DB 1368 CGTACATTATACAAAATTACATCTATTTTAAAGAAACAAATATAACGATCCAGTGT 1447
 QY 461 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 DB 1448 GGGCGCTGCTCCAAAGATTTCAGTGTCTTATCAAAAGAGAAATGCAATTACCGAT 1507
 QY 501 GlyLutTyrGluValLeuGlyLysHisGlySerAsnIleGlnValAspGlyValArgArg 520
 DB 1508 GGTGAATGGAAAGTCTTGCGCGGCAATGATTAATTCAGAGTTGATGAACTCGAAGG 1567
 QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
 DB 1568 CTGATATATTGTAAGGACACCAAGACTCCCTTTAGAGCATCACTGTACCTAGTACGT 1627
 QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
 DB 1628 TACGTAATCCCTGGAGAGGTGCAAGAGCTGACTGACCGTGGCTACTCACTTCCTGCTGC 1687
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
 DB 1688 ATCAGTCGACACTGACTGCTTTTATAGTAAGTATAGTAAACCAAGAAATCCACACTGT 1747
 QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
 DB 1748 GTGTCCCTTTACAACTATCAAGTCTCTGAGATGACCAACTGTGCAAAACAAAGGAATTT 1807
 QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspGlyTyrThrProProGluIlePhe 620
 DB 1808 TGGGCAACATTTTGGATTTCAGCAGGTCTCTCTCTGACTATACCTCCCGAATTTTTC 1867
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
 DB 1868 TCTTTTAAAGTAACTACTGATGATTTACATTTGATGGATGCTCTCAACAGCTTCATGCTA 1927
 QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
 DB 1928 CAGCTGGAAAGAAATATCTCACTGTGCTCATATATGGTGGTCTCCAGGTGCACTGTG 1987
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
 DB 1988 GTGATATATCGATTTAAAGAGTCAAGTATTTCCGCTTGATTAACCTTGACCTCTTAGGT 2047
 QY 681 TyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
 DB 2048 TATGTGTTGTGATGATGACAAACAGGGATCTCTGACCCAGGGCTTAAATTTGAAGGC 2107
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
 DB 2108 GCTTTAAATATATAATGGGTCAAAATAGAAATGACGATCGATGGAAGGACTCCAAATAT 2167
 QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
 DB 2168 CTAGCTTCTCGATATGATTTATCATTAAGTATGATGCGGATCCAGCGGTGCTCAT 2227
 QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
 DB 2228 GAGAGATACCTCTCCCTGATGGCATTAATGACAGAGTCAAGATATCTTCAAGGTTCTCAT 2287
 QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
 DB 2288 GCTGGGGGCCCAAGTACTCTGTGGATCTTCTATGATACAGGATACAGGAAAGTATATG 2347
 QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
 DB 2348 GGTACCTCGACCAAGATGAACAGGGCTATTACTTAGAGTCTGTGGCCATGCAACAGAA 2407
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 DB 2408 AAGTTCCTCTGAAACCAATCGTTACTGCTTAAACATGCTTCTCGATAGAAATGTC 2467
 QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840

DB 2468 CATTTTGACATACACAGATATATCTAGTTTTATGAGAGGCTGGAAACCATATAT 2527
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
 DB 2528 TTACAGATCTATCTCTCGAGAGACACAGCATTAAGAGTCTGTAAATCGGGAACATAT 2587
 QY 861 GluLeuHisIleLeuHisTyrLysGlnGluAsnLeuGlySerArgIleAlaIleLys 880
 DB 2588 GAATCGATCTCTTTGCACTACTACTCAAGAAACCTTGATACAGTATTCGCTCTAAA 2647
 QY 881 ValIle 882
 DB 2648 GTGATA 2653
 RESULT 4
 AX342633
 LOCUS AX342633 3106 bp DNA linear PAT 12-JAN-2002
 DEFINITION Sequence 30 from Patent WO0198468.
 ACCESSION AX342633
 VERSION AX342633.1 GI:18152030
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 AUTHORS
 1 Yue, H., Billiot, V.S., Gandhi, A.R., Lal, P., Au-Young, D.,
 Tribouley, C.M., Deleghene, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A.,
 Hafalla, A., Khan, F.A., Wajia, N.K., Yao, M.G., Lu, D.A., Peterson, C.,
 Tang, Y.T., Walsh, R.T., Azimzal, Y., Ramkumar, J., Xu, Y. and Reddy, R.
 Patent: WO 0198468-A 30 27-Dec-2001;
 JOURNAL
 Jncycle Genomics, Inc. (US)
 FEATURES
 source
 1..3106
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /note="Jncycle ID No: 7160544CB1"
 BASE COUNT 928 a 633 c 704 g 841 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 3106
 Score: 4700.00 Matches: 882
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-070-464-1 (1-882) x AX342633 (1-3106)
 QY 1 MetAlaAlaIleMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
 DB 203 ATGGCAGACGAAATGAGAAACAGACAGCTGGGTGTGGATTTGAAATCGCGACTGT 262
 QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
 DB 263 GAGAGAAATATTTGATACACAGATCGGCTTAATTTGAGCTTTTATGTTAGAGGGTAT 322
 QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
 DB 323 TCTTGAGATCGCTTAAAGCTCTTGGCCATATCCAGAAATATCATGAGCTACATGATG 382
 QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
 DB 383 GCTAAGGACACACATGATTTATATGTTGTGAAGAGATGATCCAGATGACCTCATCA 442
 QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
 DB 443 GACAGAAATCTATACCTTGCATGCTGTGAGAGAAACAGAAATATACAGCTTTTATCT 502
 QY 101 GluIleProLysThrIleAsnArgAlaValLeuMetLeuSerTyrLysProLeuLeu 120

```

Db      503 GAAATTCCTCCAAACTATCAATAGACAGACGCTCTTAATGCTCTTGGAGGCTCTTTGG 562
Qy      121 AspleuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGlnGluLeuLeuArg 140
Db      563 GATCTTTTTCAGCGACACACTGGACTATGAAATGATCTCGAAGAAAGAACTTTAAGA 622
Qy      141 GlnArgLysArgIleGlyThrValGlyTLeuAserTyrAspTyrHisGlnGlySerGly 160
Db      623 GAAAGAAAGCGCATGGACACAGTCGGAATGCTTCTTACGATTCATCCAAAGAAAGTGA 682
Qy      161 ThrPheLeuPheGlnAlaGlySerGlyTLeuTyrHisValLysAspGlyGlyProGlnGly 180
Db      683 ACATTTCTGTTTCAAGCGGATGAGATTATACGTAAAGATGAGAGGCGCCCAAGGA 742
Qy      181 PheThrGlnGlnProLeuArgProAsnLeuValGlnThrSerCysProAsnIleArgMet 200
Db      743 TTTTACGACACACTTTAAAGGCCCAATCTAGTGGAAACTAGTTGTCCTCCACATRCGATG 802
Qy      201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
Db      803 GATCCAAATTTATGCTGCTGATCCAGACTGATGCTTTTATACATACCAACGATATT 862
Qy      221 TrrIleSerAsnIleValThrArgGlnGluArgGlnThrTyrValHisAsnGlnLeu 240
Db      863 TGGATATCTACATCGTACCAAGAAAGAAAGAGACTCACTTATGTGCAAAATGACCTA 922
Qy      241 AlaAsnMetGlnGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
Db      923 GCCAACATGGAAGAGATGCGCAGATCAGCTGAGTGCTACCTTTGTTCTCCAAAGAA 982
Qy      261 PheAspArgTyrSerGlyTyrTrrPrrCysProLysAlaGlnThrThrProSerGlyGly 280
Db      983 TTTGAAATGAAATTTCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1042
Qy      281 LysIleLeuArgTrrIleLeuTyrGlnGluAsnAspGlnSerGlnValGlnIleHisVal 300
Db      1043 AAAATTTCTGAATTTCTATGTAAGAAAGAAAGATGATGAGGAGGAAATTTATCATG 1102
Qy      301 ThrSerProMetLeuGlnThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db      1103 ACATCCCTCATGTTGGAAACAGAGAGGAGCATTCATCCGTTATCTTAAACAGGTACA 1162
Qy      321 AlaAsnProLysValThrPheLysMetSerGlnIleMetIleAspAlaGlnIleArgIle 340
Db      1163 GCAAAATCCTAAGCTCTTTAAGATGTCAGAAATATGATGATGTCGAAAGAGATC 1222
Qy      341 IleAspValIleAspLysGlnLeuIleGlnProPheGlnIleLeuPheGlnGlyValGln 360
Db      1223 ATGATGTCATGATTAAGAACTAAATTCACCTTTGAGATCTATTTGAAGAGTTGAA 1282
Qy      361 TyrIleAlaArgAlaGlyTrrPrrProGlnGlyLysTyrAlaTrrPserIleLeuLeuAsp 380
Db      1283 TATATTTGCCAGACTGGATGGAATCCTCGAGGAAATATGCTTGTCATCCATCAAGAT 1342
Qy      381 ArgSerGlnThrArgLeuGlnIleValIleIleSerProGlnLeuPheIleProValGln 400
Db      1343 CGGTCCCACTGCTGCTACAGATAGTGTATCTCACCTGAATTAATTTATCCCAAGAGA 1402
Qy      401 AspAspValMetGlnArgGlnArgLeuIleGlnSerValProAspSerValThrProLeu 420
Db      1403 GATGATGTTATGGAAGGAGGAGACATCATTTGATGCTAGTCTGATCTTGAGCGCACTA 1462
Qy      421 IleIleTyrGlnGluThrThrAspIleTrrIleAsnIleHisAspIlePheHisValPhe 440
Db      1463 ATATCTATGAAGAAACACACAGCATCTGGATTAATATCATGACATCTTTATGTTTT 1522
Qy      441 ProGlnSerHisGlnGlnGlnIleGlnPheIlePheAlaSerGlnCysLysThrGlyPhe 460
Db      1523 CCCCAAGTCACGAAGAGAAATGAGTTATTTTGGCTCTGAAGCAAAACAGGTTTC 1582
Qy      461 ArgHisLeuTyrLysIleThrSerIleLeuLysGlnSerLysTyrLysArgSerSerGly 480

```

```

Db      1583 GTCATTTTATACAAATTTACATCTATTTTAAAGGAAAGCAAAATATTAACGATCCAGTGT 1642
Qy      481 GlyLeuProAlaProSerAspPheLysCysProIleLysGlnGlnIleAlaIleThrSer 500
Db      1643 GGGCTGCTCTCTCCAAATGATTTTCAAGTGTCTTATCAAAAGAGAAATGCAATTAACAGT 1702
Qy      501 GlyGlnTrrPrrValLeuGlnIleArgHisGlySerAsnIleGlnValAspGlnValArgArg 520
Db      1703 GGTGAATGGGAAGTTCCTGGCCGCGCATGATCTTATATCCAAAGTTGATGAAGTCAGAAAG 1762
Qy      521 LeuValTyrPheGlnGlnIleThrLysAspSerProLeuGlnHisHisLeuTyrValIleSer 540
Db      1763 CTGTATATTTTGAAGGACACCAAGACTCCCTTTTAAAGATACCTGTACGTAGTACGT 1822
Qy      541 TyrValAsnProGlnGlnValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db      1823 TACGTAAATCTCTGAGAGGTGACAGAGGCTACTGACCTGCTCTACTCATCTTGTGCTGC 1882
Qy      561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db      1883 ATCAGTCAGACACTGTGACTTCTTATATGATATGTAACCAAGAAATCCACACACTGT 1942
Qy      581 ValSerLeuTyrLysLeuSerSerProGlnAspAspProThrCysLysThrLysGlnPhe 600
Db      1943 GTGTCCCTTTACAGCTATCAAGTCAAGTCAAGATGACCAACTTGCAAAACAAAGAAATTT 2002
Qy      601 TrrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGlnIlePhe 620
Db      2003 TGGGCCACCAATTTGGATTTACAGAGGTCTCTTCTGACTCATCTCTCCAGAAATTTTC 2062
Qy      621 SerPheGlnSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db      2063 TCTTTTGAAGGTACTACTGATTTACATGTATGGAGTGTCTTACAGCCCTCATGATCTA 2122
Qy      641 GlnProGlnLysLysTyrProThrValLeuPheIleTyrGlnGlyProGlnValGlnLeu 660
Db      2123 CAGCTCGAAAGAAATATCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2182
Qy      661 ValAsnAsnArgPheLysGlnValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      2183 GTGAATATATGCTTTTAAAGAGTCAGATATTTCCGCTTGAAATCCCTTACGCTCTTACGT 2242
Qy      681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlnGly 700
Db      2243 TATGTGTTGTACTGATGACACAGAGGATCTCGTACCGAGGCGCTTAATTTGAAGGC 2302
Qy      701 AlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTyr 720
Db      2303 GCCTTTAAATATTAATGGTCAAAATGAGATGAGATGAGTGGAAAGACTCCAAATAT 2362
Qy      721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrrPserTyr 740
Db      2363 CTAGCTTTCGATATGATTTTCAATTCATGATGCTGATGCTGATGCTGATGCTGATGCT 2422
Qy      741 GlyGlyTyrIleSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db      2423 GGAAGATACCTCTCCCTGATGGCATTAATGACAGAGTCAGATATCTTCAAGGTTGCTATT 2482
Qy      761 AlaGlyAlaProValThrIleTrrIlePheTyrAspThrGlyTyrThrGlnArgTyrMet 780
Db      2483 GCTGGGCCCCCACTCATCTGTGATCTTCTATGATACAGGATACAGGAAACGTTATATG 2542
Qy      781 GlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGln 800
Db      2543 GGTCAACCTGACCAAGATACAGAGGCTTACTTACGATCTTGTGGCAGTCAGCAAGAA 2602
Qy      801 LysPheProSerGlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnVal 820
Db      2603 AAGTTCCTCTGAAACCAATAGTTTACTGCTTTCATGATGTTTCTCGATGAGAAATGTC 2662
Qy      821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db      2663 CATTTTGACATACACAGTATATTAATGAGTATTTTATGAGAGGCTGAAAGCCATATGAT 2722

```

QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
 Db 2723 TTACAGATCTATCCCTCAGAGAGACACAGCATTAAGAGTTCCTGATCGGAGAACATTAAT 2782
 QY 861 GluLeuHisIleLeuHisTyrLeuGlnGluSerGlySerArgIleAlaIleLeuHis 880
 Db 2783 GAACGCAATCTTTGGCACTACCTTAAGAAAACCTTGATGATCGATTGGCTCTTAA 2842
 QY 881 ValIle 882
 Db 2843 GTGATA 2848

RESULT 5
 AF221634 3127 bp mRNA linear PRI 05-NOV-2000

LOCUS Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, complete cds.
 DEFINITION AF221634
 ACCESSION AF221634.1 GI:11095187
 VERSION
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 3127)
 Abbot, C.A., Yu, D.M., Woolf, E., Sutherland, G.R., McCaughan, G.W.
 and Gorrell, M.D.
 Cloning, expression and chromosomal localization of a novel human
 dipeptidyl peptidase (DPP) IV homolog, DPP8
 Eur. J. Biochem. 267 (20), 6140-6150 (2000)

REFERENCE 2 (bases 1 to 3127)
 Abbot, C.A., Yu, D., McCaughan, G.W. and Gorrell, M.D.
 Direct Submission
 Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
 Centre, Centenary Institute of Cell Biology and Cancer Medicine,
 Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia

FEATURES
 SOURCE
 1..3127
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q22"
 1..3127
 /gene="DPP8"
 214..2862
 /gene="DPP8"
 /note="serine protease; member of prolyl oligopeptidase S9b
 family belonging to enzyme clan SC, alternatively spliced
 forms of this mRNA are found in AF221635, AF221636 and
 AF221637"

gene
 CDS
 /codon_start=1
 /product="dipeptidyl peptidase 8"
 /protein_id="AAC29766.1"
 /db_xref="GI:11095188"
 /translation="MAAMTEGUGVIFETADCEENIESODRPKLEPPYVERYSNQ
 LKILADTRKIKYMAKAPHDFFVKNPDGPHSDRIYLLAMSGNRNNTLYSEI
 PKTINRAVIMLSWPLDLFQATLDGMYREBELREKRIGTGVASIDYHOGSG
 TFLFQASGGLYHVDGPGQFTQOPLRPNVETSCPNRMPLKCPADPMIAFHEN
 DIMTINIVTREERLTYVNEHLEAMMEDASAGAVFLVEEPFRYSGYMKCAET
 PSGKILRLIYEENDESEVELIHVTSPLMTRADSRFRYKGTANPKYTFKSEIM
 DAGRIIDVIDKELIOPELIFEGVEYLIAAGWTPREKAMSILLDSQRLQIVLIS
 PELFIPVEDVMERQRLIESVPDSVPLIYEETDIDWIIHDI FHVFPQSHHELEIF
 IFASECTKFRHLTKITSILKESRYKSSGLPAPSPFKPIKEELAITGSEVEVGR
 HGSNIQVDEVARLVYEPGLOSPLNHLVYVSVNPGVETRLDRGSHSCISQHD
 FPISTYKNQKPRVSLYKLSPEPDPTCKEFWATILDSAGLPLPYTPCEIISFES
 TTGFTLYGMLYKPHDLOPGKXPTVLEITYGPOVLVNNPKYKVRRLTASIGYV
 VVIVNRGSGHRGKERGAFKYMGOIEIDQYEGLOYLASRPDIDLVGTHGMSY
 GYLISMAIMQSRSDIPRYALAGAVTITMITYDTGYTRIGHPDQNGYLLSVAMQ
 AKRPSEPRRLILHGFLELDNVAHRTSILSFLVRGKPYDQIYFQERHSIRIVES
 GEHYELHLHYLDENIGSRITALKVI"

BASE COUNT 943 a 637 c 706 g 841 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0
 Score: 4700.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 DB: 9
 Gaps: 0

US-10-070-464-1 (1-882) x AF221634 (1-3127)
 QY 1 MetAlaAlaIleMetGluThrGlnGlnLeuGlyValGluIlePheGluThrAlaASPcys 20
 Db 214 ATGCACACAGCAATGAAACAGAAACAGCTGGGTGTGAGATTTGAAACCTCGGACGT 273
 QY 21 GluGluSerIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
 Db 274 GAGGAGAAATATTGATCAAGATCGGCTTAATGAGCCTTTTATGTTGACGCGGTAT 333
 QY 41 SerTrpSerGlnLeuLysLysLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
 Db 334 TCTTGAGTCACTTAAGAAAGCTGCTTCCGATACCAAAATATCATGGCTACATGATG 393
 QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
 Db 394 GCTAAGGACACACATGATTTATCATGTTGTGAAGAGATGATCCAGATGACCTCATTC 453
 QY 81 AspArgLysTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrPheTyrSer 100
 Db 454 GACAGAACTATTACCTTGCATGTCGGAGAGAAAGAAATACAGTGTATATCT 513
 QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
 Db 514 GAATTTCCAAATCAATCAATAGACGAGCTTATGCTCTTTGGAAGCTCTTTTG 573
 QY 121 AspleuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuArg 140
 Db 574 GATCTTTTCAAGCAACATGCACTGATGATGATGATGATGATGATGATGATGATGATG 633
 QY 141 GluArgLysArgLysGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
 Db 634 GAAAGAAACGATGGAACGATGGAATGCTTCAATGATTCACCAAGAAAGGAGA 693
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
 Db 694 ACATTTCTGTTCAAGCCGATGATGATGATGATGATGATGATGATGATGATGATG 753
 QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
 Db 754 TTAAAGCAACACCTTAAAGCCCAATCTAGTGAACCTAGTTGTCACCAATACGATG 813
 QY 201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheHisSerAsnAspIle 220
 Db 814 GATTCACAAATTAAGCCCGCTGATCCAGCTGATGCTTTTAACTACCAAGAAAT 873
 QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
 Db 874 TGGATATCTAACATCGTAACAGAGAAAGAGAGATCACTATGTCACCAATGAGACTA 933
 QY 241 AlaAsnMetGluGluAsnAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
 Db 934 GCCAACATGGAAGAAAGTGCACATGACGAGTGGCTCTACCTTGTCTCCAAAGAA 993
 QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrTrpProSerGlyGly 280
 Db 994 TTGATGATATTTCTGCTATTTGATGATGATGATGATGATGATGATGATGATGATG 1053
 QY 281 LysIleLeuArgLysLeuTyrGluGluAsnAspGlnSerGluValGluIleIleHisVal 300
 Db 1054 AAAATCTTAGAATCTCATATGAAAGAAATGATGATGATGATGATGATGATGATGATG 1113

QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
 Db 1114 ACATCCCCCTATGTTGAAAACAAGAGGAGGACATTTCCTGTTATCCCTAAACAAGGTACA 1173
 QY 321 AlaAsnProLysValThrPheLysMetSerGluLeuMetIleAspAlaGluLysIle 340
 Db 1174 GCNAATCTTAAGTCACTTTTAAGATGTCAGAAATTAATGATGATGCTGAAGGAAGATC 1233
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleuPheGluIleValGlu 360
 Db 1234 ATAGATGTCATAGTAAAGAACTAATCACTTTTGAGATTCTATTGGAAGAGTTGAA 1293
 QY 361 TyrIleIleAsnArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
 Db 1294 TATATGCCAGAGCTGATGAGCTCCAGAGGAAATATGCTTGCTCATCTCACTAGAT 1353
 QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
 Db 1354 CGCTCCAGACTCGCCCTACAGATAGTGTGATCTCACCTGATTTATTTATCCAGTAA 1413
 QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
 Db 1414 GATGATGTTATGAAAAGCAGAGACCTATTGAGTCACTGCTGATTTCTGAGCCCACTA 1473
 QY 421 IleIleIleGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
 Db 1474 ATATCTATAGAAACAACAAGACATCTGATTAATTCATGATGATCATCTTCAATGTTTT 1533
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
 Db 1534 CCCCAGAGTACAGAGAGGAATGAGTTGATTTATTTTCCCTGATACCAACAGAGTTTC 1593
 QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerGly 480
 Db 1594 CGTCATTTATACAAATTAATCATCTATTTAAGGAACAATATAACATCAAGTCAGTGT 1653
 QY 481 GlyLeuProLysProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 Db 1654 GGGCTCGCTGCTCAAGTGAATTCAGAGTGCATACAAAGAGAGATAGCAATTTACAGT 1713
 QY 501 GlyIleTyrProGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
 Db 1714 GGTGAATGGGAAGTTCTTGSCCGGATGATCTTAATTCAGAGTTGATGAAAGTCAAGAG 1773
 QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleuTyrValIleSer 540
 Db 1774 CTGGTATATTTCAGAGCAACCAAGACTCCCTTTAGACATCACTGTCATGATGTCAGT 1833
 QY 541 TyrValAsnProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
 Db 1834 TACGTAATCTCTGAGAGGTGAAGAGCTGAGCTGACCGTGGCTACTACACTTCTTCTGTC 1893
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
 Db 1894 ATCAGTCAGACACTGATCTTTTATAGTAAGTAATAGTAACAGAAAGANTCCACACTGT 1953
 QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
 Db 1954 GGTGTCCTTTACAGACTATCAAGTCTGAAAGATGACCAACTGCAAAACAAGGAAATTT 2013
 QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
 Db 2014 TGGGCCACCAATTTGATTCAGAGGTCTCTTCCTGACATATACCTCTCCAGAAATTTTC 2073
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGluMetLeuTyrLysProHisAspLeu 640
 Db 2074 TCTTTGAAAGTACTACTGATTTTACATCTATAGGATGCTCTCAAGCCCTCATGATCTA 2133
 QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGluValGlnLeu 660
 Db 2134 CAGCCCTGGAAGAAATATCTACTGTGCTTCAATATATAGTGTCTCCAGGTGACAGTTG 2193
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

Db 2194 GTGAATATGCGTTTAAAGAGTCAAGTATTTCCGTTGAATACCTTAGCTCTTAGT 2253
 QY 681 TyrValValValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGly 700
 Db 2254 TATGTGGTTGATGATGATGACAAACAGGAGTCCCTGATCCGAGGAGCTTAAATTTGAAGGC 2313
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlyTyr 720
 Db 2314 GCCTTAAATATATAAAGGATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2373
 QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
 Db 2374 CTAGCTTCTGATGATGATTTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 2433
 QY 741 GlyIleTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
 Db 2434 GAGAGATACCTCTCCCTGATGACATTAATGACAGGCTCAGATATCTTCAGGGTGTATTT 2493
 QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
 Db 2494 GCTGGGCCCCAGTCACTCTGATGATTTCTTAATGATACAGATACAGGAACTTATATG 2553
 QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
 Db 2554 GGTCACTTCAAGCAGATGAAACAGGCTTATTAATGATGATGATGATGATGATGATGATGAT 2613
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 Db 2614 AAGTCCCTCTGACCAACCAATGCTTTAATGATGATGATGATGATGATGATGATGATGAT 2673
 QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
 Db 2674 CATTTGCAATACCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2733
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
 Db 2734 TTACAGATCTATCTCTGAGAGACACAGCATTAAGTCTCTGAATGCGGAGAACATATT 2793
 QY 861 GluLeuHisIleLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
 Db 2794 GAACTGATCTTTTGACATCACTTCACAAAGAAACCTTGATGACGTTATGCTGCTCTAAA 2853
 QY 881 ValIle 882
 Db 2854 GTGATA 2859
 RESULT 6
 AX354793 3143 bp DNA linear PAT 06-FEB-2002
 LOCUS AX354793
 DEFINITION Sequence 1 from Patent WO0179473.
 ACCESSION AX354793
 VERSION AX354793.1 GI:18619526
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 Meyers, R.A. and Williamson, M.
 21953, a human prolyl oligopeptidase family member and uses thereof
 Patent: WO 0179473-A 1 25-OCT-2001;
 Millennium Pharmaceuticals, Inc. (US)
 FEATURES
 source
 1. .3143
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 229..2877
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD22990.1"
 /db_xref="GI:18619527"

/translation="MAAMETBQLGVEIFETADCEINTESODRPKLEPFYVERYSWQ
LKKLADIRKYGMMAKAPHDPMFKNDPDGPHSDIYYILANGSENRENTLPFSEI
PKTINRAAVILMSKPLDLFOATLDYQWYSSEBELLEKRIIGTVAGIADYHGGSG
TFLFOAGSGIYHVKGQPGQFTQCPRLNVLVETSCPNTRMPKLCPPADIMAIHSHN
DQWISNITRREBRRLTYVHNELIANEEDASAGATVPYKPTGKNPVYTRKMSIMI
DSGGKILILYBENDESEVEILIHVTSPLERRADSPYKPTGKNPVYTRKMSIMI
DABGRIIDVIDELIOPPELLIEBVEYLARGMTEBEGYASILLDRQTLQVILIS
PELFIPEDDWNERQLIESVPDSVTPILIEETDIDINIHDIPIHVEPOSHIEIER
IFASECKTGPHILYKITSILKESKYSGGIIPASDPKCPKEKIALITSGMEVILGR
HGSNIQDVEERLIVYFEGTDSPLEHNLVYVYVNGVTRLTDRGYSGCCISQCHD
FPIKYSNOKXPHCVSLKYKLSPEDDPKCKEKPWATLIDSGPLPDYTPPEISFES
TTGPTLYGMLYKPHDLOPKKIPYTLFIYGGBOYUUNNRKGYKPYPLANTLACIYV
VVVIDNRSGSRGLKFBGAFYKMGQIISIDQVBSLQYLSRYDFTDIDRVGIGHMSY
GGYILMALMQRSDIFRVALIAGAPVLTMIFYDTGTERMYGHPDNEGGYILGSYAMQ
AEKPESEPNRLILLHGLPDENVHFAHTSILLFLVRAGKPYDLQIYQERHSIRVPS
GEHYELHLHYLOENLGSRIALAKVI"

BASE COUNT 943 a 644 c 712 g 844 t
ORIGIN

Alignment Scores:

Pred. No.: 0
Score: 4700.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Gaps: 0

Length: 3143
Matches: 882
Conservative: 0
Mismatch: 0
Indels: 0

US-10-070-464-1 (1-882) x AM354793 (1-3143)

QY 1 Metalala1a1aMetGluThrGluGluMetGluValGluIlePheGluThrAlaAspCys 20
DB 229 ATGGAGAGGAGCATGGAACAGAAACAGAGCTGGGTGTTGAGATATTTGAAACCTGGAGCTGT 288
QY 21 GluGluAsn1leGluSerGlnAspArgProIysLeuGluProPheTyrValGluIleArgTyr 40
DB 289 GAGGAGATATTTGATTCACAGGATGGCTTAATTTGAGCCCTTTTATGTTGAGCGGTAAT 348
QY 41 SerTyrSerGlnLeuIleIysLeuLeuAlaAspThrArgIysTyrHisGlyTyrMetMet 60
DB 349 TCCGTGAGTCAGCTTAAAGCTGCTGCCATACCAAGAAATATCATCGGCTACACTGTATG 408
QY 61 AlaIysAlaProHisAspPheMetPheValIysArgAsnAspProAspGlyProHisSer 80
DB 409 GCTTAGGAGCAACATGATTTTCATGTTTGTGAAGAGAAATGATCCAGATGACCTTATTC 468
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyIleAsnArgIleAsnThrLeuPheTyrSer 100
DB 469 GACAGATCTATTACCTTGCCATGCTGCTGTGAGAACAGAAATACACTGTTTATTTCT 528
QY 101 GluIleProIysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrIlePheProIleu 120
DB 529 GAAATTTCCCAAACTATCATATGAGACAGACAGCTTAAATGCTCTCTTGGAAGCCTCTTTG 588
QY 121 AspleupheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluIleuLeuArg 140
DB 589 GATCTTTTTCAGGCAACCTGACCTATGGAATGATTTCTCGAAGAAAGAACTATTTAAGA 648
QY 141 GluArgIysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
DB 649 GAAAGAAACCATTTGGAACAGTCGGAATTCCTTACGATTTATCACCAAGAAAGTGTGA 708
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValIysAspGlyValProGlnGly 180
DB 709 ACATTTCTGTTTCAAGCCGTAGTGGAAATTTATTCAGTAAAGATGAGGGCCCAAGAGA 768
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsn1leArgMet 200
DB 769 TTTAGGCAACACTTTAAGGCCCACTAGTGAACCTAGTTGCTCCCAACATACGAGATG 828
QY 201 AspProIysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
DB 829 GATCCAAATATATGCTGCTGATCCAGACTGAGATTGCTTTATATCATAGCAACAGATATT 888

QY 221 TrpIleSerAsnIleValThrArgGluGluArgIleLeuThrTyrValHisAsnGluLeu 240
DB 889 TGATATATCTAATCTGATACCAAGAGAAAGAGACTCACTTATGTGCACATATAGCTTA 948
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaIleThrPheValLeuGluGlu 260
DB 949 GCCAACATGAGAAAGAGCCAGATACAGCTGAGAGCTGACCTTGTGTTCTCCAAAGAA 1008
QY 261 PheAspArgIleSerGlyTyrTyrTrpCysProIysAlaGluThrThrProSerGlyGly 280
DB 1009 TTTGATACATTTCTGCTATTTGGTGTGTGTCGACCAAGCTGAAACACTCCAGTGTGCT 1068
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleIleVal 300
DB 1069 AAAATTTTAAATCTATATGAAAGAAATGATGATCTGAGGTGGAATATTATTCATGTT 1128
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProIysThrGlyThr 320
DB 1129 ACATCCCTTATGTTGGAACAAAGAGGCGACATTCATTCGTTATCTTAAACAGGTACA 1188
QY 321 AlaAsnProIysValThrPheIysMetSerGlnIleMetIleAspAlaGluGlyArgIle 340
DB 1189 GCMAATCTTAAGTCACCTTTTAAGATGTCAGAAATATGATTTGATCTGAAAGAAAGATC 1248
QY 341 IleAspValIleAspIysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
DB 1249 ATAGATGCTATAGATTAAGAACTAATCACTTTTGAGATTCTATTGGAAGAGATTGAA 1308
QY 361 TyrIleAlaArgAlaGlyTyrThrThrProGluGlyIysTyrAlaIleAspSerIleLeuLeu 380
DB 1309 TATATGCAAGAGCTGATGATGAGCTCTGAGGAAATATGCTGTGTCATCTTACTTAAT 1368
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
DB 1369 CGCTCCCAAGATCGCTGCAATAGTGTGATCTCACTGAAATTTTATTCACAGATAGA 1428
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
DB 1429 GATGATGTTATGAGAAAGCAGAGACTCATTTGATGACAGCTGATTCGTGACGCCACTA 1488
QY 421 IleIleTyrGluGluThrThrAspIleThrIleAsnIleHisAspIlePheHisValPhe 440
DB 1489 ATATCTTATGAGAAACCAACAGACATCTGATTAATATCCATATGATCTTTCATATGTTT 1548
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeuThrGlyPhe 460
DB 1549 CCCCAATCTACGAAGAGAAATGAGTTATTTTGGCTCTGAAATGCAAAACAGGTTTC 1608
QY 461 ArgHisLeuTyrIysIleThrSerIleLeuIysGluSerIysTyrIysArgSerSerGly 480
DB 1609 CGTCATTTATACAAATATACATCTATTTTAAAGAAAGCAATATTAACATCCAGTGGT 1668
QY 481 GlyLeuProAlaProSerAspPheIysCysProIleIysGluGluIleAlaIlePheSer 500
DB 1668 GGGCTGCTGCTCCAAAGATTTCAAGTGTCTTATCCAAAGAGAGATGCAATTAACAGT 1728
QY 501 GlyIleThrProGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
DB 1729 GGTGATGCGGAAGTTCTTGCGCGGCAATGATCTAATATCCAAAGTTGAGAGAGTCAAGAG 1788
QY 521 LeuValTyrPheGluGluThrIleAspSerProLeuGluHisHisLeuTyrValIleSer 540
DB 1789 CTGGTATATTTTGAAGGACCAAGAGACTCCCTTTAGAGCATCACTTACGTAGTGTAGT 1848
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrTyrSerHisSerCysCys 560
DB 1849 TACGTAAATCTTGAGAGAGTACAAAGGCTGACCTGACCTGCTCATCATTTCTTGCTGC 1908
QY 561 IleSerGlnHisCysAspPhePheIleSerIysTyrSerAsnGlnIysAsnProHisCys 580
DB 1909 ATCAGTGCAGACTGTGACTTCTTATATAGTAAAGTATAGTATACCGAAAGATCCACACTGT 1968
QY 581 ValSerLeuTyrIysLeuSerSerProGluAspAspProThrCysIysThrIysGluPhe 600

```

Db      1969 GTGCCCCCTTACAGCTATCAAGTCCCTAAGATGACCCAACTGCGAAACAAAGGAAATT 2028
Qy      601 TTPAlathrlleleuaspserialglyProleuproaspTyrrthProProglunlephe 620
Db      2029 TGGGCCACCATTTTGGATTGACAGAGTCTCTCTGACTATGACTCTCCCGAAGAAATTTTC 2088
Qy      621 SerhegiuserthrrthglyPhehthleuTyrglyWetleuTyrsProhisaspLeu 640
Db      2089 TCTTTGAAAGTACTACTGAGATTACATGTATGGAGGCTCTCAAGGCTTCATGATCTA 2148
Qy      641 GlnProglyLysLysTyrrProthrValleupheleTyrglyglProglunValGlnleu 660
Db      2149 CAGCTCGAAGAAAGAAATATCTCACTGCTGCTTCATATGGTGGCTCCAGGTCCAGTGTG 2208
Qy      661 ValasnansrArgPheTyrglyValLysTyrrPheargleuasnThrleuAlaSerleuGly 680
Db      2209 GTGAATATCGGTGTTAAAGGAGCAAGTATTTCCGCTTGAAATACCTAGCTCTCTAGGT 2268
Qy      681 TyrrValValValValleaspansrArgLysSerGlySarGlyLeuLysPheglunGly 700
Db      2269 TATGTGGTGTAGTAGACACAGGAGGATCTGTCAACGAGGCTTAAATTTGAAGGC 2328
Qy      701 AlaPheLysTyrrLysMetGlyGlnleuLysaspaspGlnValGlnGlyLeuGlnTyrr 720
Db      2329 GCCTTTAAATATAAATGGGTCAAAATGAAATTCAGATCAGGTGGAAGACTCCAAATAT 2388
Qy      721 LeuAlaSerArgTyrraspPheleaspLeuaspArgValGlyLeuGlyTyrrSerTyrr 740
Db      2389 CTAGCTCTCGAATAGATTTCATGACTGATGCTGTGGCCATCCAGCGCTGGTCTCTAT 2448
Qy      741 GlyGlyTyrrLeuSerleuMetAlaLeuMetGlnArgSeraspIlePheargValAlaIle 760
Db      2449 GGAGGAACTCTCTCCCTGATGCGATTAAAGCAGGCTGAGATCTTCAGGCTGCTATT 2508
Qy      761 AlaGlyAlaProValThrleuTyrrPheTyrraspThrGlyTyrrThrGlnLysTyrrMet 780
Db      2509 GCTGGGGCCCCAGTCACTGCTGGATCTTCTATGATACAGATACAGGAAAGCTTATATG 2568
Qy      781 GlyHisProaspGlnaspGlnGlnGlyTyrrTyrrLeuGlySerValAlaMetGlnAlaGln 800
Db      2569 GGTACCCCTGACAGATGAAACAGGGCTATTACTTGAATCTGTGGCCATGCAAGCAGAA 2628
Qy      801 LysPheProSerGlnProaspArgleuLeuLeuHisGlyPheLeuaspGlnAsnVal 820
Db      2629 AAGTTCCCTCTGACCAATTCGTTTACTGCTCTTACATGCTTCTGATGAGAAATGTC 2688
Qy      821 HisPheAlaHisThrSerIleleuLeuSerPheleuValArgAlaGlyLysProTyrrasp 840
Db      2689 CATTTGACATACCAAGTATATTACTGAGTTTTTGAAGGCTGGAAGGCCATATGAT 2748
Qy      841 LeuGlnleTyrrProGlnGlnLysSerIleArgValProGlnSerGlyGlnHisTyrr 860
Db      2749 TTACAAATCTATCTCAGAGAGACACAGCATAGAGTCTGAATCGGGAACAATTTAT 2808
Qy      861 GlnLeuHisleuLeuHisTyrrLeuGlnGlnLysLeuGlnSerTyrrGlnAlaAlaLeuLys 880
Db      2809 GAACCTGACTCTTTGACATGCTTCAAGAAACCTTGATCAGTATGCTGCTCTTAAAA 2868
Qy      881 ValIle 882
Db      2869 GTGATA 2874

```

REFERENCE AUTHORS TITLE JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 4535)

Strausberg, R.
Direct Submission
Submitted (27-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louised, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nantavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 71 Row: K Column: 23.
Location/Qualifiers

1. 4535
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:42570 IMAGE:4824813"
/tissue_type="Testis"
/clone_id="N1H_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
1629..4277
/codon_start=1
/product="similar to dipeptidyl peptidase 8"
/protein_id="AAH40203.1"
/db_xref="GI:26007917"
/db_xref="LocusID:54878"
/translation="MAAAMETEQLEVEIFETADCEENIESODRPKLEPFYVERYSWQ
LKTLADTRKHYGVMAKAPHDMPFVRKNDPOGPHDRYVILAMSGNRENTLPYSRI
PKTINRAVILMSKRLDLPQATLDYGMATSEEBILREKRTIGVIGASVYDHQSG
TRPQSGGTYHVDGPGQFTQDPLRPVLVETSCNIRMDPLCPADPWIAFPHGN
DIWISIVTIREERLTYVNEELANMEDBASAVAFVLOEEDRSYGWVCKAEIT
PSGGKILRIIEBENDEVEIIVHTSPMLETRADAFRYKGTSTAMPKVPKSEIMT
DAEGRILIDVDEKELIQPELIFEGVEYIARAGTTPGKXVAMNIDIRFVPSHBEIEF
PELIFVEDVMEORLIRSVDSVPLIIVERTTQIMINIDIRFVPSHBEIEF
IPASECKTGRPHLYKTTSLIKESKYSRSGGLPAPDFKPIIEETATISGEVEIEF
HGSNIVDEVRKLVIFEGTKDSPLEHLIVSVNVEVRLTDRVYSHSCCISQHD
FPISTKSNQKHPKCVSLYKLSPEDEPTCKTEFMATILDSAGLPDYPPELIFSES
TTGPTLYGMILYPHDLOPKKYPVLFVIGQVQVOLLNRRFKVYFRNLTLASLGV
VVVINDRSGCHRGKLEKGAFFKYMGOEIDDOVEGQVYASRYDFIDLBVGHGYS
GGYLSIMALMORSIDLPRAVIAAGPVLMTFVDYGVYERVWGHDPQOEGYVGSVAMO
AKPSPSEPRILLIGFDENVRPATSTILSFLVNAKRPYDQLQITPQERHSIRVES
GHHYELHLHTYQBNUGSRITAKVT"

BASE COUNT 1313 a 918 c 946 g 1358 t
ORIGIN

Alignment Scores:
Pred. No.: 0
Score: 4696.00 Length: 4535
Percent Similarity: 99.89% Matches: 881
Best Local Similarity: 99.89% Conservative: 0
Query Match: 99.91% Mismatches: 1
DB: 0 Indels: 0
Gaps: 0

US-10-070-464-1 (1-882) x BC040203 (1-4535)

```
QY      1 MetalalaIaIaMeGluThrGluGluMetGlyValGluIlePheGluThrAlaAspCys 20
Db      1629 ATGCAGACGACGATGAAACAGAACAGCTGGGTGTGAGATATTGAAACCTGGCAGCTGT 1688
QY      21 GluGluAsnIleGluSerGlnAspArgProIysLeuGluProPheTyrValGluArgTyr 40
Db      1689 GAGGGAATATTGAATCAACGAGATCGCCTTAATTGGAGCCCTTTTATGTTGAGCGGCTAT 1748
QY      41 SerTPSerGlnLeuIysLeuLeuAlaAspThrArgIysTyrHisGlyTyrMetMet 60
Db      1749 TCTTGAGATCAGCTTTAAAAGCTGCTTCCCATACCAAAATATCATGGCTACATGATG 1808
QY      61 AlaIysAlaProHisAspPheMetPheValIysArgAsnAspProAspGlyProHisSer 80
Db      1809 GCTAAGGACACACATGATTTCACTGTTTGTGAAGAGAAATGATCCAGATGACCTCATCA 1868
QY      81 AspaArgIleTyrTyrLeuAlaMeSerGlyIysAsnArgGluAsnThrLeuPheTyrSer 100
Db      1869 GACGAATCTATTTACTCTTGCCATGCTGTGTGAGAACAGAGAAATACACTGTTTATTTCT 1928
QY      101 GluIleProIysThrIleAsnArgAlaAlaValIleMetLeuSerTyrIlyAspProLeuLeu 120
Db      1929 GAAATTTCCAAAACATATCAATAGACACAGCTTAATGCTCTCTTGAGAACCTCTTTTG 1988
QY      121 AspleuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
Db      1989 GATCTTTTTCAGGCAACCTGACTATGAAATGTAATCTCGAAGAAAGAACTATTAGA 2048
QY      141 GluArgIysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db      2049 GAAAGAAACGCTTGGAAACAGTCGGAAATGCTTCTTTCAGATTATCACCAAGAAAGTGG 2108
QY      161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValIysAspGlyGlyProGlnGly 180
Db      2109 ACATTTCTGTTTCAAGCCGGTAGTGAAATTTATCATCGTAAAGATGAGGGCCACAAAGA 2168
QY      181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db      2169 TTTAGCAACACACTTTAAGGCCCAATCTAGTGAAGAACTAGTTGCCCAACATTCGGATG 2228
QY      201 AspProIysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
Db      2229 GATCCAAAATTTATGGCCCTGCTGATCCAGACTGGATGCTTTTATACATGGCAACATTT 2288
QY      221 TrpIleSerAsnIleValIleThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240
Db      2289 TGGATATCTAATCATCTGTAACAGAGAGAAAGAGACTCACTTATGTCCACATAGAGCTA 2348
QY      241 AlaAsnMetGluGluAspAlaArgSerIleGlyValAlaThrPheValLeuGlnGlu 260
Db      2349 GCCAACATGGAAGAGATGCCAGATCAGCTGAGTGGCTACCTTTGTTCTCCAGAAAGAA 2408
QY      261 PheAspArgIysSerGlyTyrTrpTyrCysProIysAlaGluThrThrProSerGlyGly 280
Db      2409 TTTGATGATATTTCTGGCTATTGGTGGTGTCCAAAAGCTGAACAACCTCCAGTGGTGT 2468
QY      281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
Db      2469 AAAATTTCTTAAGATTCATATGAAAGAAATGATGATCTGAGTGGGAAATTAATTCATGT 2528
QY      301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProIysThrGlyThr 320
Db      2529 ACAATCCCTATGTTGGAACAAAGAGGCAATTCATTCCTGTATCCCTAAAGACGGTACA 2588
QY      321 AlaAsnProIysValThrPheIysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db      2589 GCAATTCCTAAAGCTCACTTTTAAGATGTCAGAAATTAATGATTGATGCTGAAGAGATC 2648
QY      341 IleAspValIleAspIysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db      2649 ATAAATGTCATAGTAAGAACTAATTCACCTTTTGAGATTCTAATTTGAAGGAGTTGAA 2708
QY      361 TyrIleAlaArgAlaGlyTyrThrProGluGlyIysTyrAlaTyrSerIleLeuLeuAsp 380
Db      2709 TATATGGCAGAGCTGGATGATCCTCGTAGGGAATATGCTGTGGTCCATCTCATATAT 2768
QY      381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db      2769 CGCTCCGACACTCCCTACAGATAGTGTGTGATCTCACCTGGAATTAATTAATCCAGTAGAA 2828
QY      401 AspaAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db      2829 GATGATGTTATGGAAGGACAGAGACTCATAGATAGTGGCTGATTTCTGTGAGCCACATA 2888
QY      421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
Db      2889 ATTATCTATGAAAGAAACAAACAGACATCTCGATTAATATCCATGACATCTTTCACTTTT 2948
QY      441 ProGlnSerHisGlnGluGluIleGluPheIlePheAlaSerGluCysLeuThrGlyPhe 460
Db      2949 CCCCAGATCAGAGAGAGAAATGAGTTTATTTTGGCTCTGATGCAAAAACAGATTTTC 3008
QY      461 ArgHisLeuTyrIysIleThrSerIleLeuIysGluSerIysTyrIysArgSerGly 480
Db      3009 CCTCATTTATCAAAATTAATCACTTAATTTTAAAGAAACCAATATAAAGATCCAGTGGT 3068
QY      481 GlyLeuProAlaProSerAspPheIysCysProIleIysGluGluIleAlaIleThrSer 500
Db      3069 GAGCTGGCTGCTCCAGAGATTTCAAGTGTCTTCAAAAGAGAGATGCAATTACCACT 3128
QY      501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db      3129 GGTAAAGTGGAAAGTTCTTGGCCGATGAGATCTAATATCCAAAGTTGAGAAAGTCAGAAAG 3188
QY      521 LeuValTyrPheGluGlyThrIysAspSerProLeuGlnHisIleSleuTyrValIleSer 540
Db      3189 CTGGTATATTTTGAAGGACCAAAAGACTCCCTTTAGAGCATACCTGTACGTACTGACT 3248
QY      541 TyrValAsnProGlyGluValThrArgLeuThrAspArgIlyTyrSerHisSerCysCys 560
Db      3249 TACGTAAATCCTGAGAGAGTACCAAGCTGACTGACGTGGCTACTACATTTCTTGCTGC 3308
QY      561 IleSerGlnHisCysAspPhePheIleSerIysTyrSerArgGlnIysAsnProHisCys 580
Db      3309 ATCAGTACGACCTGTGACTTCTTTATAGTAAGTATGATTAACCAAGAAATCCACACGT 3368
QY      581 ValSerLeuTyrIlySleuSerSerProGluAspAspProThrCysIlyThrIlyGluPhe 600
Db      3369 GTGTCCCTTTACAAAGCATATCAAGTCCGAGATGATCCCAACTTGCAGAAACAAAGAAATT 3428
QY      601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db      3429 TGGGACACCATTTTGGATTGACGAGTGCCTTCTGACTATACCTCCAGAAATTTTC 3488
QY      621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIysProHisAspLeu 640
Db      3489 TCTTTTAAAGTACTTACTGATTTTACATTTGATGGAGTGGCTTCAAGGCTCCAGATCTA 3548
QY      641 GlnProGlyIysIysIyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db      3549 CAGCTCGAAAGAAATATTCCTACTGCTGCTGTCAATATGTTGTTCTCTCAGGTCCAGTTG 3608
QY      661 ValAsnAsnArgPheIysGlyValIysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      3609 GTGAATATCGGTTTAAAGAGTCACAGTATTTCCGTGTGAATACCTTAGGCTCTCTAGT 3668
QY      681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuIysPheGluGly 700
Db      3669 TATGTGTTGTAGTGAAGACCAAGAGGATCTCTGTACCCGAGGCTTTAATTTGAAGGC 3728
QY      701 AlaPheIysTyrIysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db      3729 GCCTTTAAATATAAATAGGCTCAAAATGAAATGACATCAGGTGGAAGGCTCCATAT 3788
QY      721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
```


QY	Db	3789	CTAGCTTCTCGAATATATTTCAATTCATGACTTAAGATCGTGTGGCATTCACGGCTGTCTAT	3848
QY		741	GIYGIYTYLLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle	760
Db		3849	GGAGGATACCTCTCCCTGATGGATTAATGACAGGTACGATATCTTCAGGGTTCTATTT	3908
QY		761	AlaGlyAlaProValThrLeuTrpIlePheTYrAspThrGlyTYrThrGluArgTYrMet	780
Db		3909	GCTGGGGGGCCCAAGTCACTCTGTGTGATCTTCTATGATACAGGATACACGGAAAGTTATATG	3966
QY		781	GlyHisProAspGlnAsnGlnGlnGlyTYrTYrTYrLeuGlySerValAlaMetGlnAlaGlu	800
Db		3969	GGTCACCCCTGACCAGATGAACAGGGCTATTTCAATGAGATCTGTGTGCCATGCACACAGAA	4028
QY		801	LysPheProSerGlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnVal	820
Db		4029	AACTTCCCTCTGAAACCAATGCTTATCTGCTCTTACATGGTTTCTCGATGATGAGATATGC	4088
QY		821	HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTYrAsp	840
Db		4089	CATTTTGGACATACCCAGTATATTAATGAGTTTATTAAGTAGGGCTGAAAGCCATATGAT	4148
QY		841	LeuGlnIleLeuTYrProGlnIluArgHisSerIleArgValProGluSerGlyGlnHisTYr	860
Db		4149	TTACAGATATATTCCTCAGAGAGAACACACATMAAGTTCCTGATCCGGAGAACATTAAT	4208
QY		861	GlyLeuHisLeuLeuHisIleTYrLeuGlnGlnIluAsnLeuGlySerArgIleAlaIleLeuHis	880
Db		4209	GAACTGCATCTTTTGACATCACTTCAAGAAACCTTGGATCAAGTATTCCTGCTCTAATAA	4268
QY		881	ValIle 882	
Db		4269	GTGATTA 4274	
RESULT 8				
AX608735				
LOCUS		AX608735	4829 bp	DNA linear PAT 17-FEB-2003
DEFINITION		Sequence 12 from Patent WO0231134.		
ACCESSION		AX608735		
VERSION		AX608735.1	GI:28404303	
KEYWORDS				
SOURCE		Homo sapiens (human)		
ORGANISM		Homo sapiens		
REFERENCE		Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS		Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.		
TITLE		Novel serine protease genes related to dppiv		
JOURNAL		Patent: WO 0231134-A 12 18-APR-2002;		
FEATURES		Ferring BV (NL)		
source		Location/Qualifiers		
		1..4829		
		/organism="Homo sapiens"		
		/mol_type="genomic DNA"		
		/db_xref="taxon:9606"		
BASE COUNT		1466 a 886 c 1017 g 1460 t		
ORIGIN				
Alignment Scores:				
Pred. No.:		0	Length:	4829
Score:		4680.00	Matches:	882
Percent Similarity:		99.77%	Conservative:	0
Best Local Similarity:		99.77%	Mismatches:	0
Query Match:		99.57%	Indels:	2
DB:		6	Gaps:	0
US-10-070-464-1 (1-882)		x	AX608735 (1-4829)	
QY		1	MetAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluTrpAlaAspCys 20	
Db		214	ATGGCAGCAGCAATGAAACAGAACAGACAGCTGGGCTTTAAGATATTTGAAACTGGCAGACTGT 273	
QY		21	GluGluAsnIleGluSerGlnAspArgProLysLeuGlnProPheTYrValGluArgTYr 40	

Db	274	GAAGAAATATGAATCAAGATCGGCTAAATTGAGCCCTTTATGTGACGGTAT	333
Qy	41	SetrSerGluLeuysLysLeuEuaIaaSPThArgLysEThiSeLysYrMetMet	60
Db	334	TCTGGAGTCAGCTTAAAAAGCTGCTGCCGATACCAAAAAATCATGTGGCTACATGAAG	393
Qy	61	AlaLysAlaProHHisAspPheMetPheValLysArgAspAspProAspGlyProHisSer	80
Db	394	GCTAAGCAACCATATTCATGTTGTTGGAAAGGAATGATCCAGATGACCTCACTTA	453
Qy	81	AspArgLysEThiLysLeuIaMetSerGlyLysArgLysGluAsnThrLeuPheYrSer	100
Db	454	GACAGATCATATTACCTTGCCATGTCGTGTGAAACAGAGAAAATACATGTTTATTC	513
Qy	101	GluLLeuProLysEThiLysAsnArgAlaValLeuMetLeuSerTrpLysProLeuLeu	120
Db	514	GAATATCCCAAAACTATCATATAGACGACGCTTAATCTCTTTGGAAGCCTCTTTTG	573
Qy	121	AspLeuPheGlnAlaThrLeuAspTrpGlyMetCysSerArgGluGluLeuLeuArg	140
Db	574	GATCTTTTTCAGGCAACACTGACACTATGGAATGTATTCGAGAAAGAAACATATTAA	633
Qy	141	GluArgLysArgLysEThiValGlyLysEThiLysEThiLysEThiLysEThiLysEThi	160
Db	634	GAAGAAACGCAATGGAACAGTCGGAATTCCTTTACGATTATCCAAAGAAATGGA	693
Qy	161	ThrPheLeuPheGlnAlaGlySerGlyLysEThiLysValLysAspGlyGlyProGlnGly	180
Db	694	ACATTTCTGTTTCAAGCCGTAAGTGAATTAATCAAGAAAGATGAGGCAACAGGA	753
Qy	181	PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnLLeuPheMet	200
Db	754	TTTAGCAACAACTTTAAAGCCCAATCTAGGAAACCTAGTGTCCCAACATACGATG	813
Qy	201	AspProLysLeuCysProAlaAspProAspTrpLysAlaPheLHisSerAsnAspLys	220
Db	814	GATCCAAATATATGCGCTGCTGATCCAGACTGATGCTTTATCATATGCAACGATAT	873
Qy	221	TrpLLeuSerAsnLLeuValThrArgLysGluLysArgLysLeuThrTrpValHisAsnGluLeu	240
Db	874	TGGAATATCTAATCATGTATCAACAGAAAGAAAGAAACCTCATATGTGCACAAATAGCTA	933
Qy	241	AlaAsnMetGluGluAspAlaArgSerLysArgValAlaThrPheValLeuGlnGluGlu	260
Db	934	GCCACATGGAAGAAAGATGCCAGATCACTGAGTGGATGCTTCTGTCTCCAAAGAA	993
Qy	261	PheAspArgTrpSerGlyTrpTrpProCysProLysAlaGluThrThrProSerGlyGly	280
Db	994	TTGTATATATATTCGCTATTCGATGGTGGTCCAAAGCTGAAACAACTCCAGTGTGT	1053
Qy	281	LysLLeuLeuArgLysLeuTrpGluLysAsnAspGlySerGluValGluLLeuLHisVal	300
Db	1054	AAAAATCTTAAATCTATATGAAGAAATATGATGATTCGAGTGAATTAATTCATGTT	1113
Qy	301	ThrSerProMetLeuGluThrArgArgLysAspSerPheArgTrpProLysThrGlyThr	320
Db	1114	ACATCCCTTACTTGGAAACAGAGGAGGCAATTCATTCGGTATCTTAAACAGATCA	1173
Qy	321	AlaAsnProLysValThrPheLysMetSerGluLysMetLLeuAspAlaGluLysArgLys	340
Db	1174	GCAATCTTAAGTCACCTTTTAAGATGCAAAATATATGATGATGCTGAAAGAAAGATC	1233
Qy	341	LLeuAspValLLeuAspLysGluLeuLLeuLProPheGluLLeuPheGluLysValGlu	360
Db	1234	ATGAATGTCATATGAAGAACTATATTCACCTTTTGATTCATTTGAAGAGATGGA	1293
Qy	361	TyrLLeuAlaArgAlaGlyTrpThrProGluLysTrpAlaTrpSerLLeuLeuAsp	380
Db	1294	TATATGGCAGAGCTGATGATCACTCCGAGGAGAAATATGCTTGGATCATCTCATAT	1353
Qy	381	ArgSerGlnThrArgLeuGlnLLeuValLeuLLeuSerProGluLeuPheLLeuProValGlu	400

Db	135	CGCTCCACAGCTCGCTACAGATAGGTGATCTACCTGAAATATTATTCOCAGTAA	1413
QY	401	AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu	420
Db	1414	GATGATGTTATGAGAAAGGACAGACCTATGATGAGTCCGATTCGTGACGCACCTA	1473
QY	421	IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe	440
Db	1474	ATTATCTATGAAAGAAACAACAACATCTGGATTAATTCATGACATCTTTCATGTTTT	1533
QY	441	ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeuThrGlyPhe	460
Db	1534	CCCCAAAGTCAGAAAGAGAAATGAGTTTATTTTGGCTCGAATGCAAAACAGTTTC	1593
QY	461	ArgHisIleuTyrLeuIlePheSerIleLeuLysGluSerLysTyrLysArgSerSerGly	480
Db	1594	CGTCATTATTAACAAATTAATCACTATTTTAAAGAAACAAATATAAACAGATCCAGTGGT	1653
QY	481	GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIlePheSer	500
Db	1654	GGGCTGCTGCTCCAAAGTGAATTCAAGTGTCTTATCAAAAGAGGATGCAATTACAGT	1713
QY	501	GlyGluTyrProGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg	520
Db	1714	GGTGATAGGAAAGTCTTGCGCGGATGATCTATATATCCAAAGTGAAGAGTCAGAAAG	1773
QY	521	LeuValIlyrPheGluGlyThrLysAspSerProLeuGlnHisIleLeuTyrValValSer	540
Db	1774	CTGGTATATTTTGAAGGACCAACAAAGACTCCCTTTAGAGCATACCTGTACGTAGTCAGT	1833
QY	541	TyrValAspProGluValValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys	560
Db	1834	TACGTAATATCCGAGAGAGTGAACAAGGTCAGTACCGTGCCTCACTCATTTCTTGCTGC	1893
QY	561	IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAspProHisCys	580
Db	1894	ATCACTACGACACTGTGACTTCTTTATAGTAAGTATACATACACAGAAATCCACACTGT	1953
QY	581	ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe	600
Db	1954	GTGTCCCTTTACAAAGCTATCAAGTCCCTAAGATGACCCAACTTGCAAAACAAAGAAATTT	2013
QY	601	TrpAlaTyrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe	620
Db	2014	TGGCCACACATTTTGGATTCACAGAGTCTCTTCTCTGACTATACCTCCCAAGAAATTTTC	2073
QY	621	SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu	640
Db	2074	TCTTTTGAAGTACTACTGATTACATTGATGGAGATGCTCTACAGCCCTCATGATCTA	2133
QY	641	GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnL	660
Db	2134	CAGCTCGAAAGAAATATCTACCTGCTGTCTTCATTAATGAGTGTCTCTCCAGGTGACGT	2193
QY	660	euvAlaAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuG	680
Db	2194	TGGGAATAAATCCGGTTTAAAGAGTCAAGATTTCCGCTTAAATACCTCAGCTCTCTAG	2253
QY	680	LysTyrValValValValIleLeuAspAsnArgLysSerCysHisArgGlyLeuLysPheGluG	700
Db	2254	GTTATGTGTGTGTGTGTGTATGACAAACAGGGGATCCTGTGCACCGAGGCTTAAATTTGAAG	2313
QY	700	LysAlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGluValGluGlyLeuGlnT	720
Db	2314	GCGCTTTAAATATTAATGGGTCAAAATAGAAATTGACGATCAGGTGAAAGACTCCAAAT	2373
QY	720	YrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyLeuHisGlyTyrPseT	740
Db	2374	ATCTAGCTTCCGATATGATATTCATGACTTAAGTATCGTGTGGGATCCACACGCGTGTCT	2433
QY	740	YrGlyGlyTyrLeuSerLeuMetLeuLeuMetGlnArgSerAspIlePheArgValAlaIat	760
Db	2434	ATGAGAGATACCTCTCTCTATGTGCATTTAATGCAAGAGGTCAAGATATCTTCAAGGTTGTCTA	2493

QY	760	leal1ad1vala1Proval1Threileu1npl1Phe1y1raa1p1n1g1y1Ty1Th1n1g1u1n1r1g1y1m	780
Db	2494	TTGCTGGGGGGCCCCAGTCACTTGTGGATCTTCTAATGATACAGGATACAGGAAAGCTTATATA	2553
QY	780	ecG1yH1s1P1r0a1S1g1n1a1n1u1n1g1y1Ty1Ty1Le1u1g1y1Se1v1A1a1w1e1c1i1n1a1g	800
Db	2554	TGGGTACACCCTCGACACAGATGAAACAGGCGTATTACTTAGATCTGTGGCCATGCAGACAG	2613
QY	800	1u1y1s1P1h1e1P1r0a1S1g1n1a1n1u1n1g1y1Ty1Le1u1g1y1Se1v1A1a1w1e1c1i1n1a1g	820
Db	2614	AAAGTTCCTCCCTCGAACCAATCGTTTACTGTCTTACATGATGTTCTTCGTGATGGAATG	2673
QY	820	a1H1s1P1h1e1A1h1i1Th1S1e1r11e1u1e1u1s1e1P1h1e1u1A1r1A1a1g1y1s1P1r01y1A	840
Db	2674	TTCAATTTTGACCAACACAGATATATATCTGATTTTATGTCAGGCGTGGAAAGCCATATG	2733
QY	840	sPl1e1u1n11e1Ty1P1r0G1n1g1u1r1g1H1s1e1r11e1a1r1v1a1P1r0G1u1s1e1r1y1G1u1H1S1T	860
Db	2734	ATTTCACAGATTCATCTTCAGAGACACACACACCA1AAAGTTCTCGAATCGGAGAACATTT	2793
QY	860	YrG1u1e1u1H1s1Le1u1e1H1s1Ty1r1e1u1g1n1u1a1n1e1u1g1y1Se1v1A1a1w1e1c1i1n1a1g	880
Db	2794	ATGACACGCACTTTTGACACTTCACTTCAAGAAACCTTGATACAGTATTCCTCTCTCTAA	2853
QY	880	YsVal1le 882	
Db	2854	AAAGTATAT 2861	
RESULT 9	BC043124	4799 bp	MRNA
LOCUS	BC043124		linear
DEFINITION	Mus musculus dipeptidylpeptidase 8, mRNA (cdna clone MGC:58057		
ACCESSION	BC043124		
VERSION	BC043124.1	GI:27695449	
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Muliyil, P., Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 4799)		
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Bluetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapellato, M., Soares, M.B., Bonaldi, M.F., Casavant, T.L., Scheer, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loughran, N.A., Peters, G.J., Abramson, R.D., Muliyil, S.J., Bosak, S.A., McEwan, P.J., McEman, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hultk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whiting, M., Madan, A., Madan, A.C., Rodrigues, S., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.I., Skalski, U., Smalinski, D.E., Schermer, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
TITLE	human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE	22388257		
REFERENCE	12477932		
AUTHORS	2 (bases 1 to 4799)		
TITLE	Strausberg, R.		
JOURNAL	Submitted (09-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		

[illegible]

OY		760	IleaaagiyalPaprovalThreLeuTriLpheTyrtasphrgiyTYrrthngiuaqTYr	779
Db		2555	ATTGCTGGAGGCCCCACGAATCACCTGTGGATCTTCTATATAACAGATACACGGAGCGGTAT	2654
OY		780	MetGlyHisProAspArgInasnGLugInglyTYrTYrLeuGlySerValAlaMetGlnAla	799
Db		2655	ATGGGTACCCTTGACCAAGAATGAACAAGGGCTACTACTTAGAATCTGTGGCCATGCACAGC	2714
OY		800	GluTyrsPheP-roserGluProAsnArgLeuleuleuHisiGlyPheLeuAspGluasn	819
Db		2715	GAGAAgTTCCCCCTGAGAACCAACCGGTACTCTCTGTGCATGGGTCTTGATGAGAAAT	2774
OY		820	ValHiSPheaIahisthrSerileLeuleuSerPheLeuValArgAlagiVlysproTYr	839
Db		2775	GTTACACTTGCACACACCATATATTCCTGA GTTTTTTGTAGAGGGCTGGAAAAGCCATAT	2834
OY		840	AspleuGlnileTYrProGingluArgHisserileArgValProGluSerGIygluHis	859
Db		2835	GACTTACAGATCTATCTCTCAGAGAGACGACAGCATCGAGTCTCGATCTGGAGAACAC	2894
OY		860	TyrGlueuhisLeuleuHisiTYrLeuGlnGluAsnLeuGlySerArgileaiAlaleu	879
Db		2885	TATGAACTGCACCTGCTCCACTACTCTTCAGAGAAACCTTGATCGGTATCGCTGCTCG	2954
OY		880	LysValille	882
Db		2955	AAAGTGATTA	2963
RESULT 10				
LOCUS		BC030688	3130 bp	mRNA linear PRI 30-MAY-2002
DEFINITION			Homo sapiens, similar to dipeptidyl peptidase 8, clone MGC:26191	
ACCESSION		BC030688		IMAGE:4822550, mRNA, complete cds.
VERSION		BC030688.1	GI:21265132	
KEYWORDS		MGC.		
SOURCE				
ORGANISM		Homo sapiens (human)		
		Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
		1 (bases 1 to 3130)		
REFERENCE		Strausberg,R.		
AUTHORS		Direct Submission		
TITLE		Submitted (24-MAY-2002) National Institutes of Health, Mammalian		
JOURNAL		Gene Collection (MGC), Cancer Genomics Office, National Cancer		
		Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
		USA		
REMARK		NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT		Contact: MGC help desk		
		Email: egapds-remail.nih.gov		
		Biusse Procurement: Miklos Palkovits, M.D., Ph.D.		
		cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki		
		Toshiyuki and Piero Carninci (RIKEN)		
		cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
		DNA Sequencing by: Institute for Systems Biology		
		http://www.systemsbio.org		
		contact: amadan@systemsbio.org		
		Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha		
		Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting		
		Clone distribution: MGC clone distribution information can be found		
		through the I.M.A.G.E. Consortium/LNL at: http://Image.lnl.gov		
		Series: IRAK Plate: 33 Row: d Column: 5		
		This clone was selected for full length sequencing because it		
		passed the following selection criteria: matched mRNA gi: 18450277.		
FEATURES		Location/Qualifiers		
source		1..3130		
		/organism="Homo sapiens"		
		/mol_type="mRNA"		
		/db_xref="taxon:9606"		
		/clone="MGC:26191 IMAGE:4822550"		
		/cissue_type="Testis"		
		/clone_lib="NIH MGC_97"		

CDS

```

/lab_host="DHN0B"
/nc="Vector: pBluescript"
372..2867
/codon_start=1
/product="Similar to dipeptidylpeptidase 8"
/protein_id="AAH30688.1"
/db_xref="gi:21265133"
/translation="MAAAMETQUGVEIFETADCEINIESODRPEFFVYRYSWQ
LKULADTRKHYGMMAAPHDPMFVKNDGDGPHSRITYLAISGERENTLFSYET
PTLNRAAVMLSMKPLDLFOALTDYDMYREBELLEKRIITGVGADYDNGSG
TFLOASGAIYHVDGQFQOPLRNLVETSOPNRMOPKICPADPDIAIHNS
DIWISNITREERLTYVHNEIAMBERASGATATVLOERPRYSYMWCPRAET
PSGGKILILYEENDESVEIHTHTSPMLERBADSRYPRKPTGIANPKVTKSEIM
DABSRITDIVDKELIOPPELIFEGEYIARAGWTPBEGYAMSILLDSQTLQVLIS
PELTPVEDDWERQRLIESVDSVTPLIYEETDIDINIHDFHVPQSEELER
IFASECKTFRHLKXITSLIKESYKRSQGLPAPDFKPIKEEIAITSGMEVLSR
HGSNIQVDEVRILYFEGTKDSPLEHLHYVSVYAPGEVRETRLDGYSHSCISQHD
FPIKSYONKPHCVSLYKLSPPEDDPCKEKEMATILDSAGPLDPYTPPEISFSS
TTGFTLGMLYKPHDLPQKKYPTVLFYIGSPQYQVNNRKYKTRFNLNLASLGY
VVYTDNGSGCHRGKKEGAPRYKVAIAGAVTILFYDGYTERYMGHPDNGGY
LGSYAMQEKFPSEPNRLILHGLDENVHAHTSILSLVRAGKEPYDILQYQERH
SIRVPSGEHYEHLHLYLOENLGSRTALAKVI"

```

BASE COUNT 955 a 626 c 717 g 832 t

Alignment Scores:

```

Pred. No.: 0 Length: 3130
Score: 4397.50 Matches: 831
Percent Similarity: 94.22% Conservative: 0
Best Local Similarity: 94.22% Mismatches: 0
Query Match: 93.56% Indels: 51
DB: 9 Gaps: 1

```

US-10-070-464-1 (1-882) x BC030688 (1-3130)

```

QY 1 MetAlaAlaIaMeGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
Db 372 ATGGCAGCAGCAATGAAACAGAACACAGCTGGTGTAGATATTGAAACGCGAGCTGT 431
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGlnProPheTyrValGluArgTyr 40
Db 432 GAGAGAGATATGATCATCAGAGATCGGCTTAATTTGAGCCTTTTATGTTGAGCGGTAT 491
QY 41 SerTyrSerGlnLeuLysLysLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db 492 TCCTGGAGTACGCTTAAAGAGCTGCTGGCATACCGAAGAAATATCATGCTCATCATGATG 551
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db 552 GCTAAGGACCAATGATGATTTCACTGTTGTGAAGAGAAATGATCCAGATGAGACCTCATCA 611
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db 612 GACAGATCTATATACCTTCCATGCTGCGTGAAGAGAAATATCATGCTTTTATTC 671
QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu 120
Db 672 GAATTCGCCAAACATTCATATAGAGAGAGAGCTTATATGCTCTCTTGAAGAGCCTCTTTTG 731
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
Db 732 GATCTTTTCAGGCAACACGACCTATGGAATGATTTCTCGAAGAAAGACATTTAAGA 791
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db 792 GAAAGAAAACGATTCGAGACAGTCGGAATGCTTCTTACGATTATCACCAAGGAAGTGGG 851
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
Db 852 AATTTCTTTTCAACCCGATGAGAAATTTATCACGTAAGAGATGAGAGGCGCAAGAGA 911
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200

```

```

Db 912 TTTAAGCAACACCTTTAAGGCCCAATCTAGTGGAATAGTGTGCCAACATACGATG 971
QY 201 AspProLysLeuCysProAlaAspProAspThrIleAlaPheIleHisSerAsnAspIle 220
Db 972 GATCCAAATATATGCGCTGATGATCCAGACCTGAGATGCTTTTATACATAGCAACATATT 1031
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
Db 1032 TGGATATCTAATCATGTAACCAAGAAAGAAAGAGACCTCATATTATGTCACATATAGCTTA 1091
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaIleThrPheValLeuGluGlu 260
Db 1092 GCCAACATGGAAGAAATGCCAGATACGCTGAGAGTCGCTACCTTGTCTCCAAAGAGA 1151
QY 261 PheAspArgTyrSerArgIleTyrTrpTyrCysProLysAlaGluThrThrProSerGly 280
Db 1152 TTATATGATATATCTGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1211
QY 281 LysIleLeuArgIleLeuTyrGluGluGluAsnAspGluSerGluValGluIleHisVal 300
Db 1212 AAAATTTCTTGAATTTCTATATGAAAGAAATATGATATGAGAGTGGAAATTTATCATGTT 1271
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db 1272 ACATCCCTTATGTTGGAACAGAGAGGAGGAGATTCATTCCTGATCTTAAACAGGTACA 1331
QY 329 AlaAspProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db 1332 GCAATATCTTAATGATGATCTTTTAAATATGATGATGATGATGATGATGATGATGATGAT 1391
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db 1392 ATAGATGTCATATGATTAAGAACTATATCAACCTTTTATGATTTATTTGAAGAGCTTGA 1451
QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
Db 1452 TATATTCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1511
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db 1512 CGCTCCAGACCTGCTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1571
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db 1572 GATGATGTTATGGAAGGAGAGACATGATGATGATGATGATGATGATGATGATGATGATGAT 1631
QY 421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
Db 1632 ATTATCTATGAAAGAAACAGACATCTGATTAATATCATGATCATCTTTCATGTTT 1691
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db 1692 CCCCAAGTACAGAAAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1751
QY 461 ArgHisLeuTyrIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
Db 1752 CGTCAATTAATACAAATATCATCTATTTAAAGAAAGCAAAATTAACATCAAGTGT 1811
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db 1812 GGGCTGCTGCTCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1871
QY 501 GlyGluTrpGluValLeuGluLysArgHisGlySerAsnIleGlnValAspGluValAlaArg 520
Db 1876 GGTGAATGGAAAGTTTGTGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1931
QY 521 LeuValTyrPheGluGluLysThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
Db 1932 CTGGATATATTTGAAGGAGCAACAAAGACTCCCTTTAAGACATCACTGACGATGACGAT 1991
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db 1992 TACGTAATCTCGAGAGGTGACAAAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2051

```

QY 561 lIeSerGlnHisCysaspPhePheIleSerIysTyrSerAsnGlnIlysAsnProHisCys 580
 Db 2052 ATCAGTCAGACACTGTACTTCTTATAGTAAGTATAGTAAACCAAGAAATCCACTGT 2111
 QY 581 ValSerLeuTyrIleuSerSerProGlnAspAspProThrCysIysThrIysGluPhe 600
 Db 2112 GATGCCCTTTACAGAGCTATCAAGTCCCTGAGAGATGACCAATGCAAAAACAAAGGAATT 2171
 QY 601 TTPAlaThrIleleuaspSerAlaGlyProLeuProAspTyrThrProGlnIlePhe 620
 Db 2172 TGGGCGACCAATTTGGATTGAGAGGTCCTCTTCCGACTATATCTCTCCAGAAATTTTC 2231
 QY 621 SerPheGlnSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIysProHisAspLeu 640
 Db 2232 TCTTTTGAAGAACTACTGATTTTACATGTATGGGATCTCTCAACAGCCTCATATCTA 2291
 QY 641 GlnProGlnIlysIysTyrProThrValLeuPheIleTyrGlyGlyProGlnIleGln 660
 Db 2292 CAGCCTGAAAGAAATATCTCTCTGCTGCTCATATATGAGTGGCTCAGGTCAGATG 2351
 QY 661 ValAsnAsnArgPheIysGlyValIysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
 Db 2352 GTGATTAATCGGTTTAAAGGAGTCAAGTATTTCCGCTGAATACCTTACCTCTAGGT 2411
 QY 681 TyrValIvalIvalIvalIleaspAsnArgGlySerCysHisArgGlyLeuIysPheGluGly 700
 Db 2412 TATGTGTTGTGATGATGATGACAAACAGGGATCTCTGACCGAGGCTTAAATTTGAAGGC 2471
 QY 701 AlaPheIysTyrIysMetGlyGlnIleGlnIleaspAsnGlnIvalGlnIleuGlnIyr 720
 Db 2472 GCCTTTAATATATAAATG----- 2489
 QY 721 LeuAlaSerArgTyrAspPheIleaspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
 Db 2489 ----- 2489
 QY 741 GlyIlyTyrIleuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
 Db 2490 -----GTTGCTATTT 2498
 QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
 Db 2499 GCTGGGGCCCGAGTACTGTGAGTCTTCTATGATACAGATACACGAAACGTTATATG 2558
 QY 781 GlyHisProaspGlnAsnGlnGlnIlyTyrTyrIleuGlySerValAlaMetGlnAlaGln 800
 Db 2559 GGTCAACCTGACCGAATGAACAGGGCTATTACTTAGCATCTGTGGCATGCAAGCAAGA 2618
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 Db 2619 AAGTTCCCTCTGAAACCAATCGTTTACTGCTCTTACATGGTTCCTGGATGAGAAATGTC 2678
 QY 821 HisPheAlaHisThrSerIleleuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
 Db 2679 CATTTTGGACATACAGATATATTACTAGTGTATTTTATAGAGGGCTGGAACCCATATGAT 2738
 QY 841 LeuGlnIleTyrProGlnGlnIleArgHisSerIleArgValProGlnSerGlyGlnHisTyr 860
 Db 2739 TTACAGATCTATCTCAGAGAGACACAGCATTAAGAGTTCCTGAATCGGAGAACATTAAT 2798
 QY 861 GluLeuHisLeuLeuHisTyrLeuGlnGlnIleuAsnIleuGlySerArgIleAlaIleLeuLys 880
 Db 2799 GAATCGCATCTTTTGCATCTACTTCAAGAAACCTTGATCAAGTATTTGCTGTCTAAA 2858
 QY 881 ValIle 882
 Db 2859 GTGATA 2864
 RESULT 11
 AX608745
 LOCUS
 DEFINITION Sequence 22 from Patent WO0231134. 4685 bp DNA linear PAT 17-FEB-2003

ACCESSION AX608745
 VERSION AX608745.1 GI:26404308
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 REFERENCE
 AUTHORS Qi, S., Akinsanya, K.O., Riviere, P.J. and Junien, J.L.
 TITLE Novel serine protease genes related to dppiv
 JOURNAL Patent: WO 0231134-A 22 18-APR-2002;
 Ferring BV (NL)
 FEATURES
 source Location/Qualifiers
 1..4685
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 1430 a 853 c 991 g 1411 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 4685
 Score: 4385.50 Matches: 834
 Percent Simlarity: 94.56% Conservative: 0
 Best Local Similarity: 94.56% Mismatches: 1
 Query Match: 93.31% Indels: 48
 DB: 6 Gaps: 1
 US-10-070-464-1 (1-882) x AX608745 (1-4685)
 QY 1 MetalAlaAlaMetGluThrGlnGlnIleuGlyValGlnIlePheGluThrAlaAspCys 20
 Db 214 ATGGCAGCAGCAATGGAAACAGAAACAGCTGGGTGTGAGATATTTGAAATCGGACCTGT 273
 QY 21 GlnGluAsnIleGlnSerGlnAspArgProIysIleuGlnProPheTyrValGlnArgTyr 40
 Db 274 GAGAGATATTTGATATCAGAGATCGGCTTAATGGAGCTTTTATGTTGAGCGGAT 333
 QY 41 SerTrpSerGlnIleuIysIysLeuLeuAlaAspThrArgIysTyrHisGlyTyrMetMet 60
 Db 334 TCTGGAGTCAAGCTTAATAAGCTCTGCGCATACCAAAATATCATGGCTACATGATG 393
 QY 61 AlaIysAlaProHisAspPheMetPheValIysArgAsnAspProAspGlyProHisSer 80
 Db 394 GCTAAGGCACACATGATTTTATGTTGTGAAGAGATGATCCAGATGACCTATTTCA 453
 QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGlnAsnArgGlnAsnThrLeuPheTyrSer 100
 Db 454 GACGAATCTATTAACCTTGCATGCTGGTGAGAAACAGAAATATACACTGTTTATCT 513
 QY 101 GlnIleProIysThrIleAsnArgAlaValIleuMetLeuSerTrpIysProIleuLeu 120
 Db 514 GAATTTCCCAAACTATCATATGACAGCAATCTTAATGCTCTCTGGAAACCTCTTTTG 573
 QY 121 AspLeuPheGlnAlaThrIleuAspTyrGlyMetTyrSerArgGlnGlnIleuLeuArg 140
 Db 574 GATCTTTTCAAGCAACATGACTAGATGAAATGATTTCTGAGAAAGAAACTTTAAGA 633
 QY 141 GluArgIysArgIleGlyThrValGlyIleAlaSerTyrAspIyrHisGlnIysGly 160
 Db 634 GAAAGAAAGCAATGGAGACAGTCGGAATGCTTTCTTACGATTTTCACAAAGGAAGTGA 693
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValIysAspGlyGlyProGlnIly 180
 Db 694 ACATTTCTGTTTCAAGCCGTAAGTGAATTTATACATTAAGATGAGGGCCCAAGGA 753
 QY 181 PheThrGlnIleProLeuArgProAsnLeuValGlnThrSerCysProAsnIleArgMet 200
 Db 754 TTACGCAACCAACTTTAAGGCCCAATCTAGTGAACCTAGTTGTCCCAATACGAGATG 813
 QY 201 AspProIysLeuIysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
 Db 814 GATCCAAATATATGAGCTGCTGATCCAGACTGGAATGCTTTTATACATACCAACGATATT 873

QY 221 TrrlleserAenllevalThrArggluGluArgleuThrTyrValHisasnGluLeu 240
 Db 874 TGGATATCTTAACATCGTAAACAGAGAAAGAGACTCACTTATGTGCACAAATGAGCTA 933
 QY 241 AlaasnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
 Db 934 GCCAACATGAGAAAGATGCCAGATCACTGAGTGGCTTGTCTTCCAAAGAA 993
 QY 261 PheAspArgIYrSerGlyTyrTrpCysProIysAlaGluThrThrProSerGlyVal 280
 Db 994 TTTGATAGATATCTGGCTATTTGGTGTGTCCAAAGCTGAACAACACTCCAGTGTGT 1053
 QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGlyIleIleHisVal 300
 Db 1054 AAAATCTCTTAAGATTCATATGAAGAAATGATGATGATGAGTGAATATTATCATGTT 1113
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgIYrProIysThrGlyThr 320
 Db 1114 ACATCCCTTATGTTGGAAACAGAGAGGAGCAATTCATTCGTTATCTTAAACAGGTACA 1173
 QY 321 AlaasnProIysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
 Db 1174 GCAATCCTTAAGTCACTTTAAGATGTCAGAAATTAATGATGATGCTGAAGAGAGATC 1233
 QY 341 IleAspValIleAspIysGluLeuIleGlnProPheGluIleLeuPheGluValGlu 360
 Db 1234 ATAAATGTCATAGATTAAGGAACTAATTCACCTTTGAGATTCATATTGGAAGAGATTGAA 1293
 QY 361 TyrIleAlaArgAlaGlyTyrTrpThrProGluGlyIYrSerTyrAlaTyrSerIleLeuLeuAsp 380
 Db 1294 TATATTCCTCAGAGTGTGATGAGACCTCTGAGGAGAAATATGCTTGTGTCATCTACATGAT 1353
 QY 381 ArgSerGlnThrArgLeuGlnIleValIleIleSerProGluLeuPheIleProValGlu 400
 Db 1354 CGCTCCAGAGCTCGCTACAGATGATGTGATCTCACCTGATATTATTCACAGTGA 1413
 QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
 Db 1414 GATGATGTATAGGAAAGGACAGACTGATGATGATGATGCTGATTTGTGACCCCACTA 1473
 QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
 Db 1474 ATATCTATGAAAGAAACACAGACATCTGATTAATATTCATGACATCTTTCATGATTTT 1533
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIYrThrGlyPhe 460
 Db 1534 CCCCAGAGTCAAGAGAGAAATGAGTTATTTTCTGCTGATCCAAACAGGTTTC 1593
 QY 461 ArgHisIleuTyrIYrIleThrSerIleLeuLysGluSerIYrIYrLysArgSerSerGly 480
 Db 1594 CGTCATTTTATACAAATTTACATCTATTTTAAAGGAAAGCAAAATATTAACGATCCAGTGT 1653
 QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 Db 1654 GGGCTCGCTGCTCCAGATGATTTCAAGTGTCTTCAAGAGAGATAGCAATTTACAGT 1713
 QY 501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGluValAspGluValArgArg 520
 Db 1714 GGTGAATGGGAAGTTCTTGCGCGCATGATCTAATATCAAGATTGATGAAGCAGAGG 1773
 QY 521 LeuValTyrPheGluGluThrLysAspSerProLeuGluHisIleuTyrValValSer 540
 Db 1774 CTGGATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACTGTAACGTAAGTCAGT 1833
 QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
 Db 1834 TAGGTAAATCTCGAAGAGGTAAGCAAGCTGACGACCGTGGCTACTCAATTTTCTGCTG 1893
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
 Db 1894 ATCAGTCAACACTGTGATCTTCTTATAGTAAATATAGTAACAGAAATCCACACTGT 1953

QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
 Db 1954 GTGTCCCTTTCAAGCTATCAAGTCCCTGAAGATGACCACTTGCAAAACAAAGCAATTT 2013
 QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
 Db 2014 TGGGCCACCATTTTGGATTTCAGT----- 2036
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
 Db 2036 ----- 2036
 QY 641 GlnProGluLysLysTyrProThrValLeuPheIleTyrGlyGlyProGluValGlnLeu 660
 Db 2037 -----CCTCAGGTGAGTTG 2051
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
 Db 2052 GTGAATATCGGTAAAGAGTCAAGATATTCGCTTGAATACCTTACCTCTAGGT 2111
 QY 681 TyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
 Db 2112 TATGTGTGTGATGATGACACACAGGAGTCTGTCAACGAGGCTTAAATTTGAAGGC 2171
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGluValGluGlyLeuGluTyr 720
 Db 2172 GCCTTAAATATATAAATGGGTCAATAGAAATTAACATACGAGTGAAGAGCTCCAAATAT 2231
 QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
 Db 2232 CTAGCTTCTGAAATGATTCATTTGATTTGATGATGATGATGATGATGATGATGATGAT 2291
 QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
 Db 2292 GAGAGATACCTCTCCCTGATGACATTATATGACAGAGTCAAGATATCTTCAAGGTTGCTAAT 2351
 QY 761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
 Db 2352 GCTGGGGCCCCAGTCACTGTGATCTTCTATGATACAGATACAGGAACTGATATATG 2411
 QY 781 GlyHisProAspGluAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
 Db 2412 GGTCAACCTTACCAAGATGAAACAGGCTATTACTTGAATCTGTGGCCATGCAAGCAGAA 2471
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 Db 2472 AAGTTCCTCTGAAACCAATCGTTTACTGTCTTACATGATGATGATGATGATGATGATG 2531
 QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
 Db 2532 CATTTGGACATACCAAGTATATTACTGAGTTTATTGAGGGCTGGAAGGCCATATGAT 2591
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
 Db 2592 TTACAGATCTATCTCAGAGACACAGCATAGAGTTCTGAATCGGAGAAACATTAT 2651
 QY 861 GluLeuHisIleuLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys 880
 Db 2652 GAACTGCATCTTTTGGACTTACCTTCAAGAAACCTTGATGATGATGATGATGATGATGAT 2711
 QY 881 ValIle 882
 Db 2712 GTGATA 2717
 RESULT 12
 AX608743 4676 bp DNA linear PAT 17-FEB-2003
 LOCUS DEFINITION Sequence 20 from Patent WO0231134.
 ACCESSION AX608743
 VERSION AX608743.1 GI:26404307
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
Ferring BV (NL)

Qi, S., Akinsanya, K.O., Riviere, P.J. and Junien, J.L.
Novel serine protease genes related to dppiv
Patent: WO 0231134-A 20 18-APR-2002;

FEATURES

Source
Location/Qualifiers
1..4676
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT
ORIGIN
1424 a 859 c 979 g 1414 t

Alignment Scores:

Pred. No.: 0 Length: 4676
Score: 4385.00 Matches: 831
Percent Similarity: 94.22% Conservative: 0
Best Local Similarity: 94.22% Mismatches: 1
Query Match: 93.30% Indels: 51
Gaps: 1

US-10-070-464-1 (1-882) x AX608743 (1-4676)

QY 1 MetlaalaalamegluThrgluGlunleuglyValgluilephegluThralaspCys 20
Db 214 ATGGAGCAGCAATGGAAACAGAGCTGGGTGTGAGATATTTGAAACCGCGACGT 273
QY 21 GluGluaenilegluSerGlnAsparProlysleuGluProphetyValgluargTyr 40
Db 274 GAGGAGAAATTTGATCATCAGAGATCGGCTTAATTGGAGCCTTTTATGTTGAGCGGTAT 333
QY 41 SerTTPSerGlnleuLysLysleuLeuAlaasphTharglystyRHisglYTYrMetMet 60
Db 334 TCCGTGAGTCAGCTTAAAGCTGCTTCCGATACCAAGAAATATCATGCTATCATGATG 393
QY 61 AlalyslaProHisaspPheMetPheValLysargAsnProaspGlyProHisSer 80
Db 394 GCTAAGGACACATGATTTTATGTTTGAAGAGAAATGATCCAGATGACCTCATTTCA 453
QY 81 AsparglileTyTYrLeuAlaMetSerGlyLysAsnargGluAsnThrLeuPheTyrSer 100
Db 454 GACAGAAATCTATTAACCTTGCATGCTGTGTGAGAAACAGAAATATCATGCTTTATTTCT 513
QY 101 GluileProlysthrileAsnargAlaAlaValleuMetleuSerTTPlysProleuLeu 120
Db 514 GAAATTCCTCAAACTATCATATAGAGACAGCTTTAATGCTCTCTGGAAGCTCTTTTG 573
QY 121 AspheupheGlnAlaThrleuAspTYrGlyMetTyrSerargGluGluLeuLeuarg 140
Db 574 GATCTTTTTCAGGCACTGTGACTATGGAATGTATTTCTCGAAGAAAGAACTATTAAGA 633
QY 141 GluargLysargileglYthValglYlleAlaSerTYrAspTYrHisglngLysSerGly 160
Db 634 GAAAGAAACGCAATGGAAAGCTGGAAATGCTCTTATCATGATTTACCAAGAAAGTGA 693
QY 161 ThrPheleupheGlnAlaGlySerGlylleTyRHisValLysaspGlyGlyProGlnGly 180
Db 694 ACATTTCTGTTTCAAGCGGTAGTGAATTTATCATCACTAAAGATGAGGCGCCCAAGGA 753
QY 181 PheThrglnGlnProleuArgProAsnleuValgluThrSerCysProAsnileArgMet 200
Db 754 TTTTACGACACACTTTAAGCCCAACTCTAGTGAATCTGTTTCCCAACATACGGAAT 813
QY 201 AspProlyseuLysProAlaaspProaspTTPlleAlaPheleHisSerAsnaspIle 220
Db 814 GATCCAAATTAATGCTGCTGATCCAGACTGATGCTTTTATACATAGAAAGATAT 873
QY 221 ThrileSerAsnileValThrargGluGluargGlyLeuThryrValHisasnGluLeu 240
Db 874 TGGATATCTAATCATGTAACAGAGAAAGAGAGACTCATTAATGTGCACAATAGACTA 933

QY 241 AlaasnMetGluGluAspAlaargSerAlaGlyValAlaThrPheValleuGlnGlu 260
Db 934 GCCAACATGAGAAAGATGCCAGATCAGCTGAGTCCGTACTTTGTTCTCCAAAGAGA 993
QY 261 PheaspArglyrSerGlyTYrTTPTrCysProlysAlaGluThrThrProSerGlyGly 280
Db 994 TTTGATAGATATTTCTGCTATTTGGTGTGTGCCAAAGCTGAACCAATCCAGTGTGT 1053
QY 281 LysileuArgGlyleuTYrGluGluAsnaspGlySerGlyValGluilleHisVal 300
Db 1054 AAAATTTCTAAGAAATTTATATGAAAGAAATATGATGTGAGGTGAAATTAATTCATGTT 1113
QY 301 ThrSerPromeLleuGluThrArgargAlaaspSerPheargTYrProlyserGlyThr 320
Db 1114 ACATCCCTTATGTTGGAACACAGAGGCAATTCATTCCTGTTATCTTAAACAGTACA 1173
QY 321 AlaasnpolyValThrPheLysMetSerGluileMetlleaspAlaGluGlyValGly 340
Db 1174 GCAATCTCTAAAGCATTTTAAAGATGTCAAGAAATATGATTTGATGCTGAAGAGATC 1233
QY 341 IleaspValleaspLysGluLeuileGlnProheGluileuPheGluGlyValGlu 360
Db 1234 ATAGATGCTATGATTAAGAACTAATTCACCTTTTGATTTCTATTTGAAGAGATTGA 1293
QY 361 TyrilleAlaargAlaGlyTTPThrProGluGlyLysTYrAlaTTPSerilleLeuLeuasp 380
Db 1294 TATATTCAGAGCTGATGATGATCTCTGAGGAAATATGCTGTGCTCATCTCTACAT 1353
QY 381 ArgSerGlnThrArgleuGlnilleValleuileSerProGluLeuPheileProValGlu 400
Db 1354 CGCTCCAGATCCCTTCAAGATGATGTGTGATTCACCTGAATTAATTAATCCAGTAA 1413
QY 401 AspaspValMetGluArgGlnArgleuileGluSerValProaspSerValThrProleu 420
Db 1414 GATAGTGTATGAAAGGACAGACTATTTAGTCACTGCTGATTCGTAGCGCACTA 1473
QY 421 IleileTyrgluGluThrThraspileTTPilleAsnilleHisaspillePheHisValPhe 440
Db 1474 ATTATCTATGAAAGAAACACAGACATCTGATTAATATTCATGATCTTTCAATGTTT 1533
QY 441 ProGlnSerHisgluGluGluileGluPheilePheAlaSerGlyCysGlyThrGlyPhe 460
Db 1534 CCCCAGAGTCAAGAGAGAAATTAATTTATTTTCCCTGTGATGCCAAAGAGGTTT 1593
QY 461 ArgHisleuTYrLysilleThrSerilleLeuLysgluSerTYrLysArgSerSerGly 480
Db 1594 CGTATTTATTAACAAATTAATCATCTATTTTAAAGAAAGCAATTAACATCCAGTGT 1653
QY 481 GlyleuProAlaProSeraspPheLysCysProilleLysgluGluilleAlaileThrSer 500
Db 1654 GGGCTGCTGCTCCAAAGTGAATTCAGTGTCTATCAAGAGAGATAGCAATTAACAGT 1713
QY 501 GluGluTTPGluValleuGlyYargHisglYserAsnilleGlnValaspGluValArg 520
Db 1714 GGTGAATGGGAAGTCTTGGCCGCGATGATCTTAATATCAAGTGTATGAAGTCAGAGG 1773
QY 521 LeuValTYrPhegluGluGlyThrLysaspSerProleuGluHisHisleuTYrValValSer 540
Db 1774 CTGGTATATTTTGAAGCAACCAAGACTCCCTTTAAGATCACTGTACGTACGTACGT 1833
QY 541 TyrValAsnProGlyGluValThrargLeuThrAsparGlyTYrSerHisSerCysCys 560
Db 1834 TACGTAATATCTGGAGGGTGAAGCAAGCTGACGACCTGTGCTAATCACTATTCCTGTC 1893
QY 561 IleSerGlnHisCysaspPhePheileSerLysTYrSerAsnGluLysAsnProHisCys 580
Db 1894 ATCAGTCAGACACTGTGACTTTTATTAAGATAGATAGTAAACCAAGAAATCCACATGT 1953
QY 581 ValSerleuTYrLysleuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db 1954 GTGTCCCTTAAAGCTATCAAGTCTGAAGATGACCAACTTGCAAAACAAAGGAATTT 2013
QY 601 TrpAlaThrilleuAspserAlaGlyProleuProaspTYrThrProGluillePhe 620

Db	714	ACATTTCTGTTTAAAGCCGCTAGTGAATTTATACGTAAAGATGAGGCGCACAGGA	773	Db	1794	CTGTATATATTTGAAAGCACCAAGACTCCCTTTAGACATACCTGTACGTACAGT	1853
Qy	181	PheThGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet	200	Qy	541	TyrValAsnProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys	560
Db	774	TTTACGCAACAACCTTTAAGGCCCATCTAGTGGAAACTAGTGTCTCCACATACGGAAG	833	Db	1854	TACGTAAATCTCGGAGAGGTGACAAAGCTGACTGACCGGTGCTACTCATCTTCTGCTGC	1913
Qy	201	AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle	220	Qy	561	IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys	580
Db	834	GATCCAAAATTATAGCCCTGCTGATCCAGACTGGAATGCTTTTATACATGCAACGATATT	893	Db	1914	ATCAGTCAGCACTGACCTCTTTTATAGTAAAGTATAGTAACCAAGAAATCCACACTGT	1973
Qy	222	TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisSerGluLeu	240	Qy	581	ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe	600
Db	894	TGGATATCTTAACCTGTAACCAAGAGAAAGAGACTCACTTATGTGCAATAGCGCTA	953	Db	1974	GTGTCCCTTTTAAAGCTATCAAGTCTCTGAAGATGACCAACTTGCAAAACAAGATTT	2033
Qy	241	AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu	260	Qy	601	TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe	620
Db	954	GCCAACTGGAAGAAAGATCCAGATCAGCTGGAGTGGCTACCTTTGTTCTCCAAAGAA	1013	Db	2034	TGGGCGACCATTTGGATTCACAGAGTCTCTCTCGATATACCTCCAGAAATTTTC	2093
Qy	261	PheAspArgTyrSerGlyTyrTyrTrpCysProLysAlaGluThrThrProSerGlyGly	280	Qy	621	SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu	640
Db	1014	TTTGATAGATATTCTGGCTATTGGTGTGCTCCAAAAGCTGAAACAACACTCCAGTGGTGT	1073	Db	2094	TCTTTTGAAGTACTACGTGAATTTACATTTATGGGATGCTCTACAAAGCTCATGATCTA	2153
Qy	281	LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal	300	Qy	641	GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu	660
Db	1074	AAAAATTTCTTAGAATTTATATAGAAAGAAATGAAATTCAGAGTGGAAATTTATCATGTT	1133	Db	2154	CAGCTGGAAAGAAATATCTCTACTGTGCTGTTCATATATGCTGCTCATG-----	2204
Qy	301	ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr	320	Qy	661	ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly	680
Db	1134	ACATCCCTCATGTGTGGAAACAAGAGGCGAGATTCATCCGTTATCTTAAACAGGTACA	1193	Db	2204	-----	2204
Qy	321	AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle	340	Qy	681	TyrValValValIleIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly	700
Db	1194	GAAAAATCTTAAAGTCACTTTTAAAGATGTCAGAAATATGATTTGATGCTGAAAGAGATC	1253	Db	2204	-----	2204
Qy	341	IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu	360	Qy	701	AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr	720
Db	1254	ATAGATGTCATGATGAAAGAACTAATTCACCTTTTGAGATTCATATTGAAAGAGTGA	1313	Db	2204	-----	2204
Qy	361	TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp	380	Qy	721	LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr	740
Db	1314	TATATTTGCCAGAGCTGATGAGTGACTCCTGAGGGAATATGCTTGATCTCATCTATAGAT	1373	Db	2204	-----	2204
Qy	381	ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu	400	Qy	741	GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle	760
Db	1374	CGCTCCAGACTCGGCTTACAGATAGTGTGATCTCACTGAAATTTTATCCCACTAGAA	1433	Db	2205	-----GTTCTATTT	2213
Qy	401	AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu	420	Qy	761	AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet	780
Db	1434	GATGATGTTATGAGAAAGCGACAGACTCATTTGAGTCAAGTGCCTGATTCGTGACGCCACTA	1493	Db	2214	GCTGGGGCCCGACACTGCTGTGATCTTATGATACAGATACACGGAACGTTATATG	2273
Qy	421	IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe	440	Qy	781	GlyHisProAspGlnAsnGluGlnGlyTyrTyrTyrLeuGlySerValAlaMetGlnAlaGlu	800
Db	1494	ATTATCTATGAAGAAACAACAGACATCTGATTAATATTCATGACATCTTTCATGTTTTT	1553	Db	2274	GGTACCTTGACCGAAATGAAACAAGGCTATTACTATAGGATCTGGGCCATGCCAAGCAAA	2333
Qy	441	ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe	460	Qy	801	LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnVal	820
Db	1554	CCCCAAATCTCAAGAGAGAAATGAGTTATTTTGGCTCTGATGCAAAACAGGTTTC	1613	Db	2334	AAGTTCCCTCTGAACCAATCGTTTACTGTCTTACATGATGTTTTCTTGAGACAGATCTC	2393
Qy	461	ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly	480	Qy	821	HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp	840
Db	1614	CGTCATTTATACAAAATTACATCTATTTTAAAGAAACAAATTAACGATCCAGTGGGT	1673	Db	2394	CATTTTCACATACACAGTAATTACTGATGTTTTTATGTAGAGGCTGGAAACCATATAT	2453
Qy	481	GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer	500	Qy	841	LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr	860
Db	1674	GGGTGCTGCTCTCAAGATGATTTCAAGTGTCTTCAAAAGAGAGATGCAATTACAGT	1733	Db	2454	TTTACAGATCTTATCTCTCAAGAGACACACAGCTAAAGATTTCTTGAAATGGGGAACATTTAT	2513
Qy	501	GlyGluTrpGluValLeuGlyValArgHisGlySerAsnIleGlnValAspGluValArgArg	520	Qy	861	GluLeuHisLeuLeuHisIstYrLeuGlnGlnLysAsnLeuGlySerArgIleAlaAlaLeuLys	880
Db	1734	GCTAAATGGGAAGTCTTTGGCCGCGATGAGATCTAATATCCAAAGTTGATGAGTAGAGG	1793	Db	2514	GAACTGCATCTTTTGACCTACCTTCAAGAAACCTTGGATCAAGATATTGCTGCTTAAAA	2573
Qy	521	LeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisHisLeuTyrValValSer	540	Qy	881	ValIle 882	
				Db	2574	GTGATA 2579	

RESULT 14
 AX608731 4523 bp DNA linear PAT 17-FEB-2003
 LOCUS Sequence 8 from Patent WO0231134.
 DEFINITION AX608731
 ACCESSION AX608731
 VERSION AX608731.1 GI:28404301
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 1 Qi, S., Akinsanya, K.O., Riviere, P.J. and Junien, J.L.
 Novel serine protease genes related to dprlv
 Patent: WO 0231134-A 8 18-APR-2002;
 Ferring BV (NL)
 FEATURES
 source Location/Qualifiers
 1..4523
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 1384 a 828 c 940 g 1371 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 4523
 Score: 4092.50 Matches: 780
 Percent Similarity: 88.44% Conservative: 0
 Best Local Similarity: 88.44% Mismatches: 1
 Query Match: 87.07% Indels: 102
 DB: Gaps: 1
 US-10-070-464-1 (1-882) x AX608731 (1-4523)
 QY 1 MetAlaAlaAlaMetGluThrGluGluGluValGluLeuPheGluThrAlaAspCys 20
 Db 214 ATGGCGAGCGACATGAAACAGAAACGCGGAGTGTGAGATATTTGAAACGCGGACGCT 273
 QY 21 GluGluAsn1LeuSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
 Db 274 GAGGAGAAATTTGAATCAAGAGTCGGCTTAATTGAGCCCTTTTATGTTGAGCGGAT 333
 QY 41 SerTyrSerGlnLeuLysLysLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
 Db 334 TCCTGAGTCAGCTTAAAGCTGTCTGCCATACCAAGAAATATCATGGCTACATGATG 393
 QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
 Db 394 GCTAAGGACACCATATATTCATGTTTGAAGAGAAATGATCCAGATCGACCTCATTC 453
 QY 81 AspArg1LeuTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
 Db 454 GACAGATCTATTAACCTTCCATGCTGCTGTGAAACAGAGAAATACACTGTTTATTC 513
 QY 101 GluLeuProLysThrLeuAsnArgAlaValLeuMetLeuSerTyrLysProLeuLeu 120
 Db 514 GAATTCCTCAAACTATCAATAGAGCAGAGCTTATATCTCTCTTGGAAGCCTCTTTG 573
 QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
 Db 574 GATCTTTTCAGGCAACATGGACATGATGATATTCGAGAGAAAGAACTATTAGA 633
 QY 141 GluArgLysArg1LeuLysThrValGlyLeuAspTyrArgTyrHisGlnLysSerGly 160
 Db 634 GAAAGAAAACGATGGAACAGTCGGAATGCTCTTACGATATATCACCAAGAAAGTGA 693
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyLeuTyrHisValLysAspGlyGlyProGlnLys 180
 Db 694 ACATTTCTGTTCAAGCCGTAAGTGAATTTATCAAGTAAAGATGAGAGGCGCAAGA 753
 QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsn1LeuArgMet 200

Db 754 TTTACGCAACACCTTTAAGGCCCAATCTAGTGAAGAACTAGTTCACCAATACGATG 813
 QY 201 AspProLysLeuCysProAlaAspProAspTyrP1LeaPhe1LeuHisSerAsnAsp1Le 220
 Db 814 GATCCAAAATATATGCTGCTGATCCAGCTGATGATGATGCTTTATATATGACCAACGATAT 873
 QY 221 Thr1LeuSerAsn1LeuVal1ThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
 Db 874 TGGATATCTAATCTGATACCAAGAGAGAGAGACTCACTTATGTGCACATGAGCTA 933
 QY 241 AlaAsnMetGluGluAspAlaArgSerArg1ValAla1ThrPheValLeuGlnGluGlu 260
 Db 934 GCCAATGAGAGAGAAATGCGACATGAGTGCAGTGCAGTGCCTTGTGCTCCAGAGAA 993
 QY 261 PheAspArgLysSerGlyTyrTyrProCysProLysAlaGluThrThrProSerArgLys 280
 Db 994 TTTGATATGATATTCCTGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1053
 QY 281 Lys1LeuValArg1LeuValArgGluGluAsnAspGluSerGluValGluLeu1LeuHisVal 300
 Db 1054 AAAATTTCTGAATTTCTATATGAAGAAATGATGATGATGATGATGATGATGATGAT 1113
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
 Db 1114 ACATCCCTATGTTGGAACAGAGAGGCGATTCATTCGTTATCTTAAACAGGTACA 1173
 QY 321 AlaAsnProLysValThrPheLysMetSerGlu1LeuMet1LeuAspAlaGluArgGly 340
 Db 1174 GCAATTCCTAAGTCACTTTAAGATGTCGAATATATGATGATGATGATGATGATGATGAT 1233
 QY 341 IleAspVal1LeuAspLysGluLeu1LeuGlnProPheGlu1LeuPheGluGlyValGlu 360
 Db 1234 ATAGATGTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1293
 QY 361 Tyr1Leu1LeuArg1LeuValProLysGluGlyTyrVal1LeuPhe1LeuLeuAsp 380
 Db 1294 TATATTCACAGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1353
 QY 381 ArgSerGlnThrArgLeuGln1Leu1Leu1LeuSerProGluLeuPhe1LeuProValGlu 400
 Db 1354 CGCTCCAGACCTGCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1413
 QY 401 AspAspValMetGluValArgGlnArgLeu1LeuGluSerValProAspSerValThrProLeu 420
 Db 1414 GATGATGTTTGAAG 1473
 QY 421 Ile1LeuTyrGluGluThrThrArgAsp1Leu1Leu1Leu1Leu1Leu1Leu1Leu1Leu 440
 Db 1474 ATTATCTATGAG 1533
 QY 441 ProGlnSerHisGluGluGlu1LeuPhe1LeuPhe1LeuPhe1LeuPhe1LeuPhe1Leu 460
 Db 1534 CCCCAGAGTCAAG 1593
 QY 461 ArgHisLeuTyrLys1LeuThrSer1LeuLeuGluSer1LeuTyrLysArgSerSerGly 480
 Db 1594 CGTATATTAACAAATATTAATCTATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1653
 QY 481 GlyLeuProAlaProSerAspPheLysCysPro1LeuLysGluGlu1LeuAla1LeuHisSer 500
 Db 1654 GGGCTGCTGCTCCTCAAGATGATTCAGAGTGCCTATCAAGAGAGAGAGAGAGAGAGAGAG 1713
 QY 501 GlyLeuThrProGluValLeuGlyArgHisGlySerAsn1LeuGlnValAspGluValArgArg 520
 Db 1714 GGTATATGAGAGAGTCTTGGCCGAGATGATCTTAATATCAAGATGATGATGATGATGATGAT 1773
 QY 521 LeuValTyrPheGluGluGlyThrLysAspSerProLeuGluHisHisLeuTyrVal1LeuSer 540
 Db 1774 CTGCTATATTTTGAAGGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1833
 QY 541 TyrValAsnProGlyGluVal1ThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
 Db 1834 TACGTAATCTCTGAG 1893

[illegible]

ACCESSION AK0000290
 VERSION AK0000290.1 GI:7020272
 KEYWORDS oligo capping, fls (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCE
 AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE
 AUTHORS 2 (bases 1 to 2778)
 TITLE Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
 JOURNAL Direct Submission
 TITLE Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1, Minato-ku Tokyo 108-8639, Japan (E-mail:cdatalim.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
 FEATURES
 source 1..2778
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HEP04088"
 /cell_line="HepG2"
 /cell_type="hepatoma"
 /clone_id="HEP"
 /note="cloning vector pME18FLJ"
 163..1572
 /note="unamed protein product"
 /codon_start=1
 /protein_id="BAA91059.1"
 /db_xref="gi:7020273"
 /translation="MAAAMEYEQGLVEIFETADCEENISODRPELPFYFERYSMGLKLTADTRKRYGMAKAPHDFFWVRNDGDGSHDRITYLAMSNGRENTLFLYSEPKTINRAAVIMLWKKPLDLFOALTDYGMYSRRELLRERICTGVASIDYQGSSTPEFASGSIYHVNDGSPQGTQQLPRLNLYETSCPNTRNDPKLCPAPDIAITHSDIYMSNVITREBRKLITYHNELANMEEDSAGAVTVLQEPFVSGYVWCPEAETTPSGKILIRILYYEENDESVEIIVHTSFMLETRADSRFYPPTGTANPKVFMKEIMDPEERIIDVIDKELIQPEFLFGEVEYIABAGMPBEKVIAMSILLDSQTLQIVLISPELRIPEADVEDMERKRLIESEVPSDFPLIYEETTDIWINIHDIHFVFPQSHBEIEIEIFASECTGFRHLIKINSI"
 BASE COUNT 856 a 562 c 615 g 745 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 2778
 Score: 4091.00 Matches: 781
 Percent Similarity: 88.55% Conservative: 0
 Best Local Similarity: 88.55% Mismatches: 1
 Query Match: 87.04% Indels: 101
 Gaps: 1
 US-10-070-464-1 (1-882) x AK000290 (1-2778)
 QY 1 MetAlaAlaAlaMetGtUthngUgnInLengUyValGhulLePhegiUthnAlaAsCys 20
 Db 163 ATGCACACGACGATGAGAAACAGAACACCTGGGGTGTAGATATTGAAACCTGCGACGTGT 222
 Yy 21 GluGluAsnLLeGluSerGlnaSPaGPrOlySLengUupProPhetyrValGluArgTyr 40
 Db 223 GAGAGAAATATTGATACACAGATCGGCTCAATTTGAGCTTTTATGTTTGGACGCTAT 282

QY	41	SerTTPSerGlnLeuLysLysLeuLeuAlaaspThrArgLysTyrHisGlyTyrIleMetCet	60
Db	283	TCCTGGAGTCAAGCTTAAAAAGCTGCTGCGATACAGAAAAATATCATGCTACATGATG	342
QY	61	AlaLysAlaProHis:AspPheMetPheValLysArgAsnAspProAspGlyProHisSer	80
Db	343	GCTAAAGGACCAACATGATTTTCATGTTTGGAAGAGAAATGATCCAGATGAGCTCATCA	402
QY	81	AspArgLysTyrTyrLeuAlaMetSerGlyLysAsnArgLysGlnAsnThrPhePheTyrSer	100
Db	403	GACGAACTCTATTACTTCTGCGCATGCTGCTGGTGAACAGAGAAAAATACACTGTTTATCT	462
QY	101	GluLysPhePheLysThrLysAsnArgLysAlaValLeuMetLeuSerTyrLysPheLeuLeu	120
Db	463	GAATTTCCCAAAACATATCAATAGACACAGCTTAAATGCTCTTGGAGAGCTCTTTTG	522
QY	121	AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg	140
Db	523	GATCTTTTTCAGGCAACCTGACTATGAAATGTAATTCGAGAAAGAAACTATTAGA	582
QY	141	GluArgLysArgLysGlyThrValGlyLysLeuAspTyrAspTyrHisGlyLysGly	160
Db	583	GAAGAAAGAACCATTTGGAACAGCTGCGAATTCCTTTACGATTATACCAAGAAAGTGA	642
QY	161	ThrPheLeuPheGlnAlaGlySerGlyLysTyrHisValLysAspGlyGlyProGlnGly	180
Db	643	ACATTTCTGTTTCAAGCCGTAAGTGAATTATACGTAAGAAAGATGAGGAGCCACAAGGA	702
QY	181	PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnLysArgMet	200
Db	703	TTTACGCAACACCTTTTAAAGCCCAATTTAGTGAAGAACTAGTGTCCCAACATCGAGATG	762
QY	201	AspProLysLeuCysProAlaAspProAspTyrLysAlaPheLysSerAsnAspLys	220
Db	763	GATCCAAATATTATGCTCCCTGCTGATCCAGACTCGAATGCTTTTATACATAGCAAGATTT	822
QY	221	TyrLysSerLysLysValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu	240
Db	823	TGGATATCTTAACATCGTAACCGAGAAAGAAAGAGACTCTTATGTCACAAATGACGTA	882
QY	241	AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu	260
Db	883	GCCAACTGGAAGAGATGCCAGATCCAGCTGAGAGTCCGTACTTGTGTTCCAGAAAGAA	942
QY	261	PheAspArgTyrSerGlyTyrTyrTyrCysProLysAlaGluThrThrProSerGlyGly	280
Db	943	TTTGATAGATATTCTGCTATTGCTGATGTCGCAAAAGCTGAACAACCTCCAGAGTGCT	1002
QY	281	LysLysLeuArgLysLeuTyrGluGluAsnAspGluSerGluValGluLysLysVal	300
Db	1003	AAAATTTCTTGAATTTCTATATGAAAGAAATGATGAACTGAGTGGAATTAATTCATGTT	1062
QY	301	ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr	320
Db	1063	ACATCCCTCTATTGTTGAAACAGAGAGGCGCATTCATTCGTTATCTTAAACAGATACA	1122
QY	321	AlaAsnProLysValThrPheLysMetSerGluLysMetLysAspAlaGluLysArgLys	340
Db	1123	GCAAAATCTTAAAGTCACTTTTAAGATGTCAGAAATATATGATGCTGGAAGAGAGATC	1182
QY	341	LysAspValLysLysLysGluLeuLysLysGlnProPheGluLysLeuPheGluLysVal	360
Db	1183	ATMGATGTCATATAGATAGAACTAATTCACCTTTTGAAGATCTATTTTGAAGAGATTGA	1242
QY	361	TyrLysAlaArgAlaGlyTyrThrProGluGlyLysTyrValLysPheSerLysLeuLeuAsp	380
Db	1243	TATATGCGCAGAGCTGAGTGAATCTCTGAGGAGAAATATGCTTGCTCATCTACATAGAT	1302
QY	381	ArgSerGlnThrArgLeuGlnLysLysLysSerProGluLeuPheLysProValGlu	400
Db	1303	CGCTCCCAAGACTGCTACAGATAGATGTTGATCTCACTGAATTAATTAATCCAGATGAA	1362

QY	401	AspAspValMetGluArgGlnArgLeuLysGlnSerValProAspSerValThrProLeu	420
Db	1363	GATGATGTTATGGAAGAGCAGAGACTCATTTGATGATAGTGCTGATTTCTGTGAGCCACTA	1422
QY	421	LysLysTyrGluGluThrThrAspLysLysLysLysLysLysLysLysLysLysLysLys	440
Db	1423	ATTATCTATGAAAGAAACAGACATCTGATTAATATCCATGATCATCTTTCATGTTT	1482
QY	441	ProGlnSerHisGluGluGluLysGluPheLysPheAlaSerGluCysLysThrGlyPhe	460
Db	1483	CCCCAAAGTCAAGAGAGAAATGATTTATTTTGGCTCTGAAATGCAAAACAGGTTTC	1542
QY	461	ArgHisLysTyrLysLysLysSerLysLysLysGlnSerLysTyrLysArgSerGly	480
Db	1543	CGTCAATTATACAAATATTAATTAAT-TTAAAGAAACCAATATTAAGATCCAGATGCT	1601
QY	481	GlyLeuProAlaProSerAspPheLysCysProLysGlnGluLysAlaAlaLysSer	500
Db	1602	GAGCTGCTGCTCCAGATGATTTCAGAGTCTCTATCCAAAGAGATAGCAATTCAGCT	1661
QY	501	GlyGluTyrGluValLeuGlyArgHisGlySerAsnLysGlnValAspGluValArgArg	520
Db	1662	GGTAAATGGGAAGTCTTGCCGCGCATGATCTAATATCCAAAGTTGATGAAGTCAGAAAG	1721
QY	521	LeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisLysLeuTyrValLysSer	540
Db	1722	CTGCTATATTTTGAAGGACCAAGACCTCCCTTTAGAGCATCACTTATGATCTATGCTAGT	1781
QY	541	TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys	560
Db	1782	TACGTAAATCTGAGAGAGTGACAAAGCTGATGACCTGCTGCTACTCATTTCTTGCTGC	1841
QY	561	LysSerGlnHisCysAspPhePheLysSerLysTyrSerAsnGlnLysAsnProHisCys	580
Db	1842	ATCAGTACAGACTGTGACTTCTTTATAGTAAGTAATGTAACCAAGAAATCCACACACTG	1901
QY	581	ValSerLeuTyrLysLysSerSerProGluAspAspProThrCysLysThrLysGluPhe	600
Db	1902	GTCCTCCCTTTCAGAGCTATCAAGTCCGTAAGATGACCCAACTTGGAAGAAACAGAAATTT	1961
QY	601	TyrAlaThrLysLeuAspSerAlaGlyProLeuProAspTyrThrProProGluLysPhe	620
Db	1962	TGGGCCACCAATTTGATGATTCAGAGAGTCTCTTCTGACTATCTCTCCAGAAATTTTC	2021
QY	621	SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu	640
Db	2022	TCCTTTGAAAGTACTACTGATTTTACATTTGATGAGATGCTCTCAAGCCATCATGATCTA	2081
QY	641	GlnProGlyLysLysTyrProThrValLeuPheLysTyrGlyGlyProGlnValGlnLeu	660
Db	2082	CAGCTGGAAGAAATATCTCTACTGCTGCTGCTCATATATGATGCTGCTCAG-----	2132
QY	661	ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly	680
Db	2132	-----	2132
QY	681	TyrValValValValLysAsnArgLysSerCysHisArgGlyLeuLysPheGluGly	700
Db	2132	-----	2132
QY	701	AlaPheLysTyrLysMetGlyGlnLysGluLysAspGlnValGluGlyLeuGlnTyr	720
Db	2132	-----	2132
QY	721	LeuAlaSerArgTyrAspPheLysAspLeuAspArgValGlyLysHisGlyTyrSerTyr	740
Db	2132	-----	2132
QY	741	GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspLysPheArgValAlaLys	760
Db	2133	-----	2133
QY	761	AlaGlyAlaProValThrLeuTyrLysPheTyrAspThrGlyTyrThrGluArgTyrMet	780

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:14:56 ; Search time 648.054 Seconds

(without alignments)
3673.932 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700

Sequence: 1 MAAMMETRQLGVEIFETADG.....HLHYLDENLGSRIALAKVI 882

Scoring table:

BLOSUM62	
Xgapop 10.0, Xgapext 0.5	
Ygapop 10.0, Ygapext 0.5	
Fgapop 6.0, Fgapext 7.0	
Delop 6.0, Delext 7.0	

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10070464/rumat.15102003.113552.24799/app.query.fasta_1.2652
-DB=N-Geneseg.19Jun03 -OPMT=fastcap -STUFFIX=ring -MINMATCH=0.1 -LOOFCI=0
-LIST=45 -DOCLIG=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US10070464.@CGN.1.1.0.@rnatc.15102003.113552.24799 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELACK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N-Geneseg.19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4700	100.0	2649	25	ABX12255	cDNA encoding huma
2	4700	100.0	2671	24	ABK83322	cDNA encoding huma
3	4700	100.0	3106	24	ABK12892	Human protease PRT
4	4700	100.0	3120	22	AAK85694	Nucleotide sequenc
5	4700	100.0	3120	24	AAK38956	Human dipeptidyl p
6	4700	100.0	3143	24	AAH99934	cDNA encoding 2195
7	4695	99.9	2643	24	AAH99935	Coding sequence of
8	4680	99.6	4829	24	ABK83327	cDNA encoding huma
9	4385.5	93.3	4685	24	ABK83332	cDNA encoding huma
10	4385	93.3	4676	24	ABK83331	cDNA encoding huma
11	4118	87.6	2842	24	ABN59774	Novel human coding
12	4092.5	87.1	4523	24	ABK83325	cDNA encoding huma
13	3970.5	84.5	2510	24	AAK23843	Human protease PRT
14	3771	80.2	2668	24	ABN59775	Novel human coding
15	3661.5	77.9	4309	24	ABK83328	cDNA encoding huma
16	3364.5	71.6	2161	22	AAH15009	Human cDNA sequenc
17	2870	61.1	2617	24	ABK83323	Human cDNA sequenc
18	2870	61.1	3716	24	ABO75955	Human PMMW encodin
19	2870	61.1	4219	24	ABK83335	Human cDNA encoding
20	2870	61.1	4302	24	ABK83333	cDNA encoding huma
21	2863	60.9	3024	24	AAK38954	Human dipeptidyl p
22	2835	60.3	2495	24	AAK38957	Human dipeptidyl p
23	2833	60.3	3287	24	AAK38955	Alternative versio
24	2820.5	60.0	4180	24	ABK83339	cDNA encoding huma
25	2820.5	60.0	4263	24	ABK83338	cDNA encoding huma
26	2801	59.6	1821	24	ABV76411	Dipeptidyl peptida
27	2763	58.8	2751	24	AAK83311	Murine dipeptidyl
28	2649	56.4	4076	24	ABK83337	cDNA encoding huma
29	2649	56.4	4159	24	ABK83336	cDNA encoding huma
30	2638	56.1	2801	22	AAI57896	Human polynucleoti
31	2599.5	55.3	4037	24	ABK83341	cDNA encoding huma
32	2599.5	55.3	4120	24	ABK83340	cDNA encoding huma
33	2476.5	52.7	3262	22	AAK157880	Human polynucleoti
34	2422	51.5	1669	22	AAK65696	Nucleotide sequenc
35	2226.5	47.4	2982	24	AAI59666	Human polynucleoti
36	1914.5	40.7	2461	21	AAK57835	Human ORFX ORF1390
37	1878.5	40.0	2952	24	ABK69030	DNA encoding human
38	1874.5	39.9	3047	24	ABK69113	DNA encoding human
39	1836.5	39.1	1083	22	AAK85687	Nucleotide sequenc
40	1645.5	35.0	1197	22	AAK85695	Nucleotide sequenc
41	1644.5	35.0	2027	21	AAK77137	Human ORFX ORF2692
42	1599.5	34.0	3733	23	ABL10425	Drosophila melanog
43	1599.5	34.0	3783	23	ABL06641	Human peptidase, H
44	1400	29.8	2079	21	AAA37672	Human cDNA #843 d1
45	1400	29.8	2079	25	ABK63843	

ALIGNMENTS

RESULT 1

ABX12255 standard; cDNA; 2649 BP.

ABX12255;

19-MAY-2003 (first entry)

cDNA encoding human serine protease HIPHM46.

KW Human; ss; gene; HIPHM46; serine protease; gene therapy; osteoarthritis;
KW serine protease activity; modulation; dipeptidyl peptidase activity;
KW musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
KW Alzheimer's disease; parasympathetic palsy; Huntington's disease;
KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
KW irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;

KM colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
 KM multiple sclerosis; chromosome 15q21-q22.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Key 1..2649
 XX CDS /*tag= a
 XX FT /product= "HIPHM46"
 XX FT /note= "Serine protease"
 XX
 XX GB2374869-A.
 XX
 XX 30-OCT-2002.
 XX
 XX 22-JAN-2002; 2002GB-0001404.
 XX
 XX 23-JAN-2001; 2001GB-0001760.
 XX
 XX (GLAXO) GLAXO GROUP LTD.
 XX
 XX Edbrooke MR, Lewis AP;
 XX WPI; 2003-150703/15.
 XX P-PSDB; AB007720.
 XX
 XX Identifying modulators of serine protease activity useful for treating
 XX muscular diseases, by contacting cell expressing a novel serine
 XX protease polypeptide with a compound and monitoring serine protease
 XX activity -
 XX
 XX Claim 12; Page 22-26; 38pp; English.
 XX
 XX The invention relates to a method of identifying a substance that
 XX modulates serine protease activity, comprising contacting a cell such as
 XX a neuronal cell, lung cell, intestinal cell or a cell infected with a
 XX virus, expressing a serine protease polypeptide (HIPHM 46), or its
 XX variant having dipeptidyl peptidase activity, or a serine protease
 XX isolated from the cell with a test substance and monitoring for serine
 XX protease activity. The method is useful for identifying a substance that
 XX modulates serine protease activity. A modulator of the serine protease is
 XX useful in the manufacture of a medicament for treatment or prophylaxis of
 XX a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
 XX infection, Alzheimer's disease, paraspinal nuclear palsy, myotonic
 XX dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
 XX Additional disease that may be treated using modulators of the serine
 XX protease include malabsorption syndromes, irritable bowel syndrome, lung
 XX disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
 XX rectal polyps, small bowel tumours, colorectal tumours, anaemia,
 XX dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
 XX sclerosis. The present sequence represents cDNA encoding the human serine
 XX protease HIPHM46 which is located on chromosome 15q21-q22.
 XX
 XX SQ Sequence 2649 BP; 803 A; 514 C; 585 G; 747 T; 0 other;
 XX
 XX Alignment Scores:
 XX Pred. No.: 0 Length: 2649
 XX Score: 4700.00 Matches: 882
 XX Percent Similarity: 100.00% Conservative: 0
 XX Best Local Similarity: 100.00% Mismatches: 0
 XX Query Match: 100.00% Indels: 0
 XX Db: 25 Gaps: 0
 XX
 XX US-10-070-464-1 (1-882) x ABX12255 (1-2649)
 QY 1 MetAlAlAlAMetGluThrGluGluLeuGlyValAluIlePheGluThrAlaAspCys 20
 Db 1 ATGGCAGCGCATGGAAACAGACAGCTGGGTTGAGATATTGAAACCTGGGACTGT 60
 QY 21 GluGluAsnIleGluSerGlnAspArgProlGlySleuGluProPheTyrValGluArgTyr 40
 Db 61 GAGGAAATATATGATACACAGATCGCTTAATGGACCTTTTATGTGAGCGGAT 120

QY 41 SerTTPSerGluLeuGlySleuLeuAlaAspThrArgGlySerTyrHisGlyTyrMetMet 60
 Db 121 TCTTGAGTCAAGCTTAAAGCTGCTTGCAGATACAGAAATATCATGCTACATGATG 180
 QY 61 AlaAlaProHisAspPheMetPheValIleArgAsnAspProAspGlyProHisSer 80
 Db 181 GCTAAGGACCAACATGATTTTCATGTTTGCAAGAGAAATATCCAGATGAGACCTCATCA 240
 QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
 Db 241 GACAGATCTATTACCTTGCCATGTCTGTGAGAAACAGAAATAATACATGTTTATTTCT 300
 QY 101 GluIleProlGlyThrIleAsnArgAlaAlaValIleMetLeuSerTyrPylsProlLeu 120
 Db 301 GAAATCCCAAACTATCAATAGAGCAGCTTAATGCTTCTTGAAACCTCTTTTG 360
 QY 121 AspleupheGlnAlaThrIleAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
 Db 361 GATCTTTTTCAGGCACACCTGACTATGATGATGATCTCGAAGAAAGAACTTTAAGA 420
 QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlySerGly 160
 Db 421 GAAGAAACCGCATTTGAAACAGTGGAAATGCTTACATATTCACCAAGAAAGTGA 480
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValIleAspGlyGlyProlGly 180
 Db 481 ACATTTCTGTTTCAAGCCGAGTGAATTTATACGTAAGAAATGAGAGGCCACAAAGA 540
 QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
 Db 541 TTTACGCAACACTTTAAGGCCCAATCTAGTGAATCTAGTTCCCAATACGAGATG 600
 QY 201 AspProlGlySerCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
 Db 601 GATCCAAATATATTCCTCTGCTGATCCAGACGAGTGTGTTTAAACATAAGAACCAATAT 660
 QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
 Db 661 TGGATATCTAACATCATGTAACAGAGAAAGAAAGAACTCATTTATGTCACAAATGAGCTA 720
 QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValIleuGluGlu 260
 Db 721 GCCAATCATGAAAGAAAGATCCAGATCAGCTGAGTCCGCTACTTGTCTCCAAAGAA 780
 QY 261 PheAspArgTyrSerGlyTyrTyrTyrCysProlGlyAlaGluThrThrProlSerGly 280
 Db 781 TTTGATATATATTTCTGGCTATTTGGTGTGTCCTCAAAAGCTGAACAACTCCAGTGTGT 840
 QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
 Db 841 AAAATCTTAAGATTTCTATGAAAGAAATATGATGATCTGAGGTGAAATTTATCATGTT 900
 QY 301 ThrSerProlMetLeuGluThrArgArgAlaAspSerPheArgTyrProlGlyThr 320
 Db 901 ACATCCCTATGTTGGAACAAAGAGGCGCAGATTCATTCCTTATCTTAAACAGGTACA 960
 QY 321 AlaAsnProGlyValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
 Db 961 GCAAAATCTTAAGTACATTTTAAAGATGTCAGAAATAATGATGATGATGAGAGAGATC 1020
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
 Db 1021 ATGATGTCATATGATTAAGAACTAATTCACCTTTTGAGATTTCTTTGAAGAAGGTGA 1080
 QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyTyrValAlaTyrSerIleLeuLeuAsp 380
 Db 1081 TATATTCGACAGAGCTGATGATGATCTCTGAGGAAATATATGCTGTGCTTACTATGAT 1140
 QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
 Db 1141 CGCTCCAGACTCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProl 420

CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinestias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.

XX Sequence 2671 BP; 805 A; 524 C; 594 G; 748 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	2671
Score:	4700.00	Matches:	882
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-070-464-1 (1-882) x ABK83322 (1-2671)

```

QY 1 MetAlaAlaMetGluThrGluGlnLeuGlyValGluLeuPheGluThrAlaAspCys 20
DB 8 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGAGATATTGAACTGGCGACTGT 67
QY 21 GluGluAsnLleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
DB 68 GAGAGAGATATGATCATCAGATCGGCTTAAATGGAGCTTTTATGTTGACCGGTAT 127
QY 41 SerTyrSerGlnLeuLysLysLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
DB 128 TCTGTGAGTCAAGCTTAAAGCTGCTGCGATACCAAGAAATATCATGCTCATCATGATG 187
QY 61 AlaValAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
DB 188 GCTAAGGACCAATGATGATTTTCACTTTGTAAGAGAAATATCCAGTGGACCTCATTTCA 247
QY 81 AspArgLleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
DB 248 GACGAGATCTATTAACCTTGGCATGCTGAGAGAAACAGAAATACACTGTTTATTTC 307
QY 101 GluLleProLysThrLleAsnArgAlaAlaValLeuMetLeuSerTrrLysProLeuLeu 120
DB 308 GAAATTCCTCAAACTATCAATAGAGAGACAGTCTTAATGCTCTTGGAAAGCCTCTTTTG 367
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
DB 368 GATCTTTTCAAGCAACACTGGACCTATGGAATGATTTCTCGAAGAAAGAACTATTAAAG 427
QY 141 GluArgLysArgLleGlyThrValGlyLleAlaSerTyrAspTyrHisGlnGlySerGly 160
DB 428 GAAAGAAACGCAATGGAAACAGTCGAATGCTTTACGATTATACCAAGCAAGATGGA 487
QY 161 ThrPheLeuPheGlnAlaGlySerGlyLleTyrHisValLysAspGlyGlyProGlnGly 180
DB 488 ACATTTCTGTTTCAACCGGTAGTGAATTTATCACTAAAGATGGAAGGCGCAAGGA 547
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnLleArgMet 200
DB 548 TTTAGCAACAACCTTTAAGGCCCAATCTAGTGAAGAACTAGTTGTCCTCAACATACG 607
QY 201 AspProLysLeuCysProAlaAspProAspTrrLleAlaPheLleHisSerAsnAspLle 220
DB 608 GATCCAAATATATGCCCTGCTGATCCAGACTGATTTGCTTTATACATAGCAACGATATT 667
QY 221 TrpLleSerAsnLleValThrArgGluGluArgArgLeuThrTyrValHisGlnGluLeu 240
DB 668 TGGATATCTAATCATGCTAATCCAGAGAAAGAAAGAGACTACTATATGTGACATAGCTCA 727
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
DB 728 GCCAACATGGAAGAGATCCAGATCAGCTGAGAGTCTGACTTGTCTCCAAAGAGAA 787

```

```

QY 261 PheAspArgTyrSerGlyTyrTrrLysProLysAlaGluThrThrProSerGlyGly 280
DB 788 TTTGATGAGATATTCGCTATGAGTGTGCTCAAAAGCTAAACAACTCCAGGTGGTGT 847
QY 281 LysLleLeuArgLleLeuTyrGluGluAsnAspGlySerGluValGluLleLleHisVal 300
DB 848 AAAATCTTATGAAATTCATATGAAAGAAATGAAATCTAGGTGGAATTTATTCATGTT 907
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
DB 908 ACATCCCTATGTTGGAACCAAGAGGCGAGATTCATCTGTTATCCTTAACAACGATACA 967
QY 321 AlaAspProLysValThrPheLysMetSerGluLleMetLleAspAlaGluGlyArgLle 340
DB 968 GCAATCTCAAACTCACTTTAAGATGTCAGAAATATATGATGCTGAAGAGAGATC 1027
QY 341 IleAspValLleAspLysGluLeuLleGlnProPheGluLleLeuPheGluGlyValGlu 360
DB 1028 ATGATGTCATATGATAGAACTAATTCACCTTTGAGATTTCTATTGAAAGAGTTGAA 1087
QY 361 TyrLleAlaArgAlaGlyTrrThrProGlnGlyLysTyrAlaTrrSerLleLeuLeuAsp 380
DB 1088 TATATTGCCAGAGCTGATGCACTCTGAGGGAATATGCTTGTCATCTCATCTGAT 1147
QY 381 ArgSerGlnThrArgLeuGlnLleValLleLleSerProGluLeuPheLleProValGlu 400
DB 1148 CGCTCCAGACTGCTGCGAGATGAGTGTGATTCACCTGAATATTTATCCACAGTAGAA 1207
QY 401 AspAspValMetGluArgLleGluSerValProAspSerValThrProLeu 420
DB 1208 GATGATGTTATGGAAGGACAGACCATTTAGTGTGCTGATTTCTGAGCGCCACTA 1267
QY 421 LleLleTyrGluGluThrThrAspLleTrrLleAsnLleHisAspLlePheHisValPhe 440
DB 1268 ATTATCTAAGAAACCAACAGACATCTGATTAATATCATGATCTTTCACTGTTT 1327
QY 441 ProGlnSerHisGluGluGluLleGluPheLlePheLleSerGluCysLysTrrGlyPhe 460
DB 1328 CCCCAAGTCAAGAGAGAAATTTGATTTATTTTGGCTCTGAAAGCAAAAGAGTTTC 1387
QY 461 ArgHisLeuTyrLysLleThrSerLleLeuLysGluSerLysTyrLysArgSerGly 480
DB 1388 CCGTATTTATCAAAATTAATCATCTATTAAAGAAAGCAATATTAACATCACTGAT 1447
QY 481 GlyLeuProAlaProSerAspPheLysCysProLleLysGluGluLleAlaLleThrSer 500
DB 1448 GGGCTGCTGCTCCCAAGTATTCAGTCTCTATCAAGAGAGATGCAATTAACAGT 1507
QY 501 GlyLutrrGluValLeuGlyArgHisGlySerAsnLleGluValAspGluValArgArg 520
DB 1508 GGTGAATGGGAAGTCTTGCGCGCATGATCTAATATCAAGTTGATGAGACAGAG 1567
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValLysSer 540
DB 1568 CTGGATATTTTGAAGGACCAAGACTCCCTTTAAGACATCACTGATAGTACAT 1627
QY 541 TyrValAspProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
DB 1628 TACGTAATCTCGGAAGGTGCAAGCTGATGACCGGTGCTACTCACAATCTTGCTG 1687
QY 561 IleSerGlnHisCysAspPhePheLleSerLysTyrSerAsnGlnLysAsnProHisCys 580
DB 1688 ATCACTGACGACATGACATCTTTATTAAGTAAATGATTAAGTAAACAGAGATCACTGT 1747
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
DB 1748 GTGTCCCTTTCAACACTATCAAGTCTGAAGATGACCCAACTTGCAAAAGAAATTT 1807
QY 601 TrpAlaTrrLleLeuAspSerAlaGlyProLeuProAspTrrThrProProGluLlePhe 620
DB 1808 TGGGCCACATTTGGATTGACAGAGTCTCTCTGACTATATCTCCACGAAATTTTC 1867

```

```

OY 621 SerpneGluSerThrThrglyPheThrLeuTyrGlyMetLeuTyrIleProHisAspLeu 640
DB 1868 TCTTTGAAAGTACTCTGATTTACATTGTATGGAGATGCTCTACAAAGCTCATGATCTA 1927
OY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
DB 1928 CAGCTGGAAGAAATATCTCTACTGCTGCTTATATGAGTGTCTCTCAGGCGAGCTTG 1987
OY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAsnLeuGly 680
DB 1988 GTGAATTAATCGATTAAAGAGTCAAGTATTTCCGCTTAATACCTTACCTCTAGGT 2047
OY 681 TyrValValValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGly 700
DB 2048 TATGCTGTTGATGATGATTAACAACAGGAGATCTGTCCACCGAGGCTTAAATTTGAAGGC 2107
OY 701 AlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnLysLeuGlnTyr 720
DB 2108 GCCTTTAATATTAATAATGAGTCAAAATAGAAATTTGACGATCAGAGTGGAAAGACTCCAAAT 2167
OY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleGlyTyrPheTyr 740
DB 2168 CTAGCTTCTCGATGATGATTTCACTTACATTGATCGTGGGATCCACGCGTGTGCTCAT 2227
OY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
DB 2228 GGAGGATACCTCTCCCTGATGAGCATTTATGCAAGGCTCAGATATCTTCAAGGCTGCTCAT 2287
OY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlnArgTyrMet 780
DB 2288 GCTGGGCGCCAGTCACTCTGAGATCTTCTATGATACAGGATACAGGAACTGTTATG 2347
OY 781 GlyHisProAsnGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
DB 2348 GGTACCTGTCACCAATGAACAGGGCTATTACTTAGATCTGTGGCCATGCAACAGCAA 2407
OY 801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsnGlnAsnVal 820
DB 2408 AAGTTCCCTCTGAAACCAATCGTTTACTGCTCTTACATGGTTCTCGAAGAGATGTC 2467
OY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
DB 2468 CATTTGGACATACCAAGTATTAATCTAGTGTATTTAGAGAGGGCGGAAACCATATAT 2527
OY 841 LeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGlnSerGlyLysIleTyr 860
DB 2528 TTACAGATCTATCTCTCAGAGAGACACAGCATTAAGATTTCTGAACTCGAGAAACATTAT 2587
OY 861 GlnLeuHisLeuLeuHisTyrLeuGlnGlnLeuAsnLeuGlySerArgGlyIleAlaLeuLys 880
DB 2588 GAATCGATCTTTTGGCACTACCTTCAGAAACCTTGATCAGATATGCTGCTTAA 2647
OY 881 ValIle 882
DB 2648 GTGATA 2653

```

RESULT 3
ABK12892
ID ABK12892 standard, cDNA, 3106 BP.

AC ABK12892;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-9 cDNA sequence.
XX
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis; ss.
XX

```

OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 203..2851
FT /tag= a
FT /product= "Human protease PRTS-9"
PN WO200198468-A2.
PD 27-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-US19178.
XX
PR 16-JUN-2000; 2000US-212336P.
XX
PR 22-JUN-2000; 2000US-213955P.
XX
PR 29-JUN-2000; 2000US-215396P.
XX
PR 07-JUL-2000; 2000US-216821P.
XX
PR 14-JUL-2000; 2000US-218946P.
XX
PA (INCYTE) INCYTE GENOMICS INC.
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM,
PI Delegeane AM, Baughn MR, Nguyen DB, Lee EA, Hatfield A, Khan FA;
PI Walla NK, Yao WG, Lu DM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallik DA;
XX
DR WPI; 2002-090437/12.
DR P-PsDB; AAU74749.
XX
PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX
PS Claim 5; Page 166-167; 177pp; English.
XX
CC The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present nucleic acid sequence encodes
CC the human protease PRTS-9 protein of the invention.
XX
SQ Sequence 3106 BP; 928 A; 633 C; 704 G; 841 T; 0 other;

```

Alignment Scores:

Pred. No.: 0
Score: 4700.00 Length: 3106
Percent Similarity: 100.00% Matches: 882
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-10-070-464-1 (1-882) x ABK12892 (1-3106)

```

OY 1 MetAlaAlaAlaMetGluThrGlnGlnLeuGlyValGlnIlePheGluThrAlaAspCys 20
DB 203 ATGGACACACCAATGGAAACAGACAGCTGGTTCAGATATTGAAACCTGCGCAGACTGT 262
OY 21 GlnGlnAsnIleGluSerGlnAspArgProLysLeuGlnubProPheTyrValGlnArgTyr 40
DB 263 GAGGAGATATTTGATACACGATCGCCCTTAATTTGAGAGCTTTTATGTTGACGGGAT 322
OY 41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60

```

323 TCCTGAGTACAGCTTAAAGCTGCTGCCATACAGAAAATATCATGCTACATGATG 382
QY 61 AAlaValaProHisPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
DB 383 GCTAAGGACACACATGATTCATGTTTGGAAGAGAAATGATCCAGATGAGCTCATCA 442
QY 81 AspArgIleTyrTyrIleuAlaMetSerGlyLysAsnArgGluAsnThrLeuPheTyrSer 100
DB 443 GACAGATCTATTACTCTGCTGCTGCTGAGAAACAGAAAATATCATGTTTATCT 502
QY 101 GlnIleProLysThrIleAsnArgAlaValLeuMetLeuSerTyrLysProLeuLeu 120
DB 503 GAAATCCCAAACTATCATAGAGACAGACTTAATGCTCTCTTGAGAGCTCTTTTG 562
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGlnGluLeuMetArg 140
DB 563 GATCTTTTCAGCAACACTGAGCTATGATGATGATTCGAGAAAGAACTATTAAGA 622
QY 141 GluArgLysArgIleGlyThrValGlyIleAsnTyrAspTyrHisGlnGlySerGly 160
DB 623 GAAAGAAAACGATTTGGAACAGTCGGAATTCCTTTACGATTATCACCAAGAAAGTGA 682
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
DB 683 ACATTTCTGTTTCAGACCGGTAGTGAATTTATCAAGTAAAGATGAGGCGCACAGAGA 742
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 743 TTTCAGCAACACTTTAAGGCCCAATCTAGTGAACCTAGTTGTCACCAACTACAGATG 802
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
DB 803 GATCCAAATATATGCTCTGCTGATCCAGCTGATTCCTTTATACATAGCAACATATTT 862
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
DB 863 TGATATCTTACATGCTATACAGAGAGAGAAAGAGAGCTCATATGTCACATAGACTA 922
QY 241 AlaAsnMetGluGlnAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
DB 923 GCCAATGAGAAAGAAATGCGAGATCAGCTGAGAGTCCGCTACTTGTCTCCAAAGAAA 982
QY 261 PheAspArgTyrSerGlyTyrTyrTrpCysProLysAlaGluThrThrProSerGlyGly 280
DB 983 TTGATAGATATTTCTGGCTATGCTGCTGCTCCAAAGCTGAAACAACCTCCAGTGTGT 1042
QY 281 LysIleLeuArgIleLeuTyrGlnGluLysAsnAspGluSerGluValGluIleHisVal 300
DB 1043 AAATATCTTAGAATTCATATAGAAAGAAATGATGATGAGTGGAAATTTATCATGTT 1102
QY 301 ThrSerProMetLeuGlnThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
DB 1103 ACATCCCTATGTTGAAACAGAGAGGAGGAGATCTCTCGTTATCTTAAACACAGTACA 1162
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
DB 1163 GCAATCTTAAGTCACTTTTAAGATGTCAGAAATATATGATTGATGCTGAGAGAGATC 1222
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGlnIleLeuPheGlnGlyValGlu 360
DB 1223 ATAGAGTCAATAGATAGAACTAATTAACCTTTGAGATTCCTTTGAGAGAGTGTAA 1282
QY 361 TyrIleAlaArgAlaGlyTyrThrProGlnGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
DB 1283 TATATGCGAGAGCTGATGAGACTCTGAGGAGAAATATGCTTGCTCATCTCATATGAT 1342
QY 381 ArgSerGlnThrArgLeuGlnIleValIleuIleSerProGlnLeuPheIleProValGlu 400
DB 1343 CGCTCCCAAGCTGCTCAAGATAGTGTGATCTCACTGAAATTTTATATCCAGATAGA 1402
QY 401 AspAspValMetGluLysGlnArgLeuIleGlnSerValProAspSerValThrProLeu 420
DB 1403 GATGATGTTATGGAAGAGAGAGACTCATTTGAGTCACTGCTGATCTGTGAGCGCACTA 1462

QY 421 IleIleTyrGlnGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
DB 1463 ATATCTATGAAAGAAACAAACAGACATCTGAAATTAATATCAAGACATCTTATGTTT 1522
QY 441 ProGlnSerHisGlnGluGlnIleGlnPheIlePheAlaSerGlnCysLysThrGlyPhe 460
DB 1523 CCCCAGATGACAGAGAGAGAAATGAGTTATTTTGGCTCTGCAATGCAAAACAGTTTC 1582
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGlnSerLysTyrLysArgSerGly 480
DB 1583 CGTCAATTTATACAAATTTACATCTATTTAAAGAAAGCAAAATATTAACATCCAGTGTG 1642
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGlnGlnIleAlaIleThrSer 500
DB 1643 GAGCTGCTGCTCAGATGATTTCAAGTGTCTTCAAGAGAGAGATGCAATTTACAGT 1702
QY 501 GlyLeuTrpGlnValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
DB 1703 GGTGAATGGAAGATTCCTGCGCGCATGATCTATATATCCAGTTGATGAAAGTCAAGAG 1762
QY 521 LeuValTyrPheGlnGlyThrLysAspSerProLeuGlnHisHisLeuTyrValValSer 540
DB 1763 CTGATATATTTTAAGGACAAAGACTCCCTTTAGAGATCACTGATGATGATGATGAT 1822
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
DB 1823 TAGGTAAATCTCGAGAGAGTGAACAGAGCTGATGACCGTGGCTATCCACTTCTGCTGC 1882
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
DB 1883 ATCAGTACAGACGTGACTTTTATTAAGTAAATATGATTAACAGAGAAATCCACTGT 1942
QY 581 ValSerLeuTyrLysLeuSerSerProGlnLysAspProThrCysLysThrLysGluPhe 600
DB 1943 GTGTCCCTTACAAAGCATTAATCAATCTCGAAGATGACCACTTGCAAAACAAAGAAATT 2002
QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
DB 2003 TGGGCCACCAATTTGATTCAGACAGGCTCTTCTGACTATATCTCTCCAGAAATTTTTC 2062
QY 621 SerPheGlnSerThrThrArgLysPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
DB 2063 TCTTTTAAAGTACTACTGATTTTACATTTATGAGAGTGTCTTCAAGGCTCATGATCTA 2122
QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyLysProGlnValGlnLeu 660
DB 2123 CAGCTGGAAGAAATATCTCTATGCTGCTCATATATGATGATGATGATGATGATGATG 2182
QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
DB 2183 GTGATATATGATTTAAAGAGTCAAGTATTTCCGCTGAAATCCCTAGCCTCTCATGAT 2242
QY 681 TyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlnGly 700
DB 2243 TATGTGTTTATGATGATAGAGCAACAGGAGATCTGTACACGAGGCTTAAATTTGAAGC 2302
QY 701 AlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnValGlnGlyLeuGlnTyr 720
DB 2303 GCCTTTAAATATTAAGAGGTCAAAATGAATGAAGATCAGGTGAGAGAGATCCCAATAT 2362
QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
DB 2363 CTAGCTCTGATATGATTTTCACTTGAATGATGATGATGATGATGATGATGATGATGAT 2422
QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
DB 2423 GAGAGATACCTCTCCCGATGAGATTAATGACAGAGTCAAGATATCTTCAAGGATGCTAT 2482
QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlnArgTyrMet 780
DB 2483 GCTGGGCCCCAGTCACTGTGATCTTATGATACAGATACAGAAAGCTTATATG 2542

QY 761 GLYHISProAspGluAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
 DB 2543 GGTACCCCTGACAGATGAACAGGCTATTACTTAGATCTGTGGCCATGCAAGCAGAA 2602
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 DB 2603 AAGTTCCTCCCTGACCAAAATGTTTACTGCTTACATGATGTTTCTTGATGAGAAATGTC 2662
 QY 821 HISPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlnAlaGlyLysProTyrAsp 840
 DB 2663 CATTTTGACATACACAGATATATTAATGAGTTTATGAGAGGCTGGAAGCCATATGAT 2722
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyr 860
 DB 2723 TTACAGATCTATCCCTCAGAGAGACACACAGCATTAAGATTCTGATGGGAGAACTATTAT 2782
 QY 861 GlnLeuHisLeuLeuHisTyrLeuGlnGlnLeuGlnGlySerArgIleAlaIleLeuLys 880
 DB 2783 GAACCTGACATCTTTTGCACCTACCTTCAGAAAACCTTGATCAGATATGCTCTTAAA 2842
 QY 881 ValIle 882
 DB 2843 GTGATA 2848

RESULT 4
 AAC85694
 ID AAC85694 standard; cDNA; 3120 BP.
 XX AAC85694;

AC AAC85694;
 XX 29-JUN-2001 (first entry)

DE Nucleotide sequence of human DPP8.

XX Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
 KM dipeptidyl peptidase; DPPIV; T cell; cleavage: diatriosea;
 KM growth hormone deficiency; glucose level; mucosal regeneration;
 KM non-insulin dependent diabetes mellitus; glucose intolerance;
 XX immunosuppression; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 214..2862
 FT /*tag= a
 FT /product= "Human DPP8"

XX WO200119866-A1.

XX 22-MAR-2001.

XX 11-SEP-2000; 2000MO-AU01085.

XX 10-SEP-1999; 99AU-0002762.

XX 18-FEB-2000; 2000AU-0005709.

XX (UNSY) UNTV SYDNEY.

XX Abbot CA, Gorell MD;

XX WPI; 2001-281520/29.

XX P-PSDB; AAB47187.

XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
 PT substrates, identifying inhibitors of DPP8 catalytic activity which
 PT have therapeutic uses, and for detecting activated T cells

XX Claim 16; Fig 2; 78pp; English.

XX This sequence encodes human dipeptidyl aminopeptidase (DPP8). DPP8
 CC has substrate specificity for H-Ala-Pro-PNA, H-Gly-Pro-PNA and
 CC H-Arg-Pro-PNA. Therefore, it is a prolyl oligopeptidase and a
 CC dipeptidyl peptidase, because it is capable of hydrolysing the

CC peptide bond C-terminal to Pro in each of these compounds. DPP8
 CC is homologous with human DPPIV. DPP8 is useful for cleaving a
 CC substrate, and for detecting an activated T cell which involves
 CC measuring the level of DPP8 gene expression in a T cell. The level
 CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
 CC in the cell. It is also useful for identifying a molecule capable
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules
 CC identified as inhibiting DPP8 catalytic activity may be useful for
 CC treating diabetes, growth hormone deficiency, lowering glucose levels
 CC in non-insulin dependent diabetes mellitus and other disorders
 CC involving glucose intolerance, enhancing mucosal regeneration and
 CC as immunosuppressants.

XX Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:
 Pred. No.: 0 Length: 3120
 Score: 4700.00 Matches: 882
 Percent Similarity: 100.00% Conservative: 0
 Best local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-10-070-464-1 (1-882) x AAC85694 (1-3120)

QY 1 MetAlaAlaMetGluThrGlnGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
 DB 214 ATGGCAGCAGCATATGAAACAGACAGCTGGGCTTGAGATATTGAACTGGGACTGT 273
 QY 21 GlnGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
 DB 274 GAGAGAAATTTTAAATCAGACAGATCGCTTAAATTGAGCGCTTTTATGTGAGCGGTAT 333
 QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
 DB 334 TCTCGAGTCAGCTTTAAAAGCTGCTTCCGACACAGAAAATATATGCTCATGATG 393
 QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
 DB 394 GCTAAGCAGCACATGATATTCATGTTTGCAAGAGAAATGATCAGATGAGACTCATTTCA 453
 QY 81 AspArgGlyLeuTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
 DB 454 GACAGATCTATTAATCTTCCATGCTGCGTGAAGAAACAGAAATACACTGTTTATTC 513
 QY 101 GlnIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu 120
 DB 514 GAATTCGCCAAACTTCATATAGCAGACAGAGCTTATATGCTCTTGGAAAGCCTCTTTG 573
 QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGlnLeuLeuArg 140
 DB 574 GATCTTTTTCAGGCAACACTGAGCTATGGAATGTATTCGAGAAAGAACTATTAAAG 633
 QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
 DB 634 GAAAGAAAAGCATTTGGAACAGTCGAAATGCTCTTACGATTATACCAAGAGATGGA 693
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyLysProGlnGly 180
 DB 694 ACATTTCTGTTTCAACCGGTAGTGAATTTATCATCGTAAAGATGAGAGGCCACAAAGA 753
 QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
 DB 754 TTTAACGAAACACTTTAAGGCCCAATCTAGAGAAACATGATGTCCTCAACATACGAGAG 813
 QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
 DB 814 GATCCAAAATTAATAGCCCGCTGATCCAGACTGGAATGCTTTTATATACATGCAACGATAT 873
 QY 221 TrpIleSerAsnIleValThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240
 DB 874 TGGATATCTAAACATCTGTAACAGAGAAAGAGACTCATATGTGACATATGAGCTTA 933

241 AlaAsmMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluGlu 260
 934 GCCAACATGGAGAAAGANTGCCAGATCAGCTGGAGTGCATCACTTTGTTCTCCAAAGAGAA 993
 261 PheAspArgTyrSerGlyTyrTrpTyrCysProValAlaGluThrThrProSerGlyGly 280
 994 TTGTAAGATGATTTCTGGCTATTTGGTGGTCTCAAAAGCTGAAACAACTCCCAAGTGGT 1053
 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValAlaGluIleIleVal 300
 1054 AAAATCTTGATTCATCATATGAAGAAATGATGAAATCTGAGGTGGAAATATATCATGTT 1113
 301 ThrSerProMetLeuGluThrArgAlaAlaAspSerPheArgTyrProLysThrGlyThr 320
 1114 ACATCCCTATGTGGAAACAAAGAGGCGATTCATTCGTTATCTTAAACAGATACA 1173
 321 AlaAspProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
 1174 GCAAACTCTAAAGTCACTTTTAAGATCTCGAAATTAATGATGATGCTGAAGAGAGATC 1233
 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluValGlu 360
 1234 ATGATGTCATAGATVMAAGAACTATTCACCTTTGAGATTCATTTGAAGAGATTGAA 1293
 361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
 1294 TATATGCCAGAGCTGATGAGACTCCGAGGAGAAATATGCTGGTCACTCACTAGAT 1353
 381 ArgSerGlnThrArgLeuGluIleValLeuIleSerProGluLeuPheIleProValGlu 400
 1354 CGCTCCAGACTGCCCTACAGATAGTGTGATCTCACTGAATTAATTTATCCAGTGA 1413
 401 AspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProLeu 420
 1414 GATGATGTATGAAAGGAGAGACTCATTCAGTCACTGCTGATTCGAGACCCCACTA 1473
 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
 1474 ATATATCATGAAGAAACAACAGACATCTGATTAATTCATGACATCTTTCATGTTTTT 1533
 441 ProGlnSerHisGluGluGluIleGluPheIlePheIleAspGluCysLysThrGlyPhe 460
 1534 CCCCAAGTCCAGAAAGGAAATTCAGTTATTTTTTGGCTGTGATGCAAAACAGGTTTC 1593
 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
 1594 CGTCATTTATACAAATATACATCTATTTAAAGAAACAAATATTAACGATCCAGTGGT 1653
 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 1654 GGGGCTGCTGCCAAGATGATTTCAAGTGTCTTATCAAAAGAGAGATTCGCAATTCACGT 1713
 501 GlyLeuTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
 1714 GGTGAATGGAAAGTTCTTGGCCGGCATGATCTAATATCCAAAGTTGAGATCGAAGG 1773
 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
 1774 CTGGATATTTTGAAGGACCAAAACCTCCCTTTAAGACATCACTGTACGTACGTACGT 1833
 541 TyrValAsnProGluGluValThrArgPheThrAspArgGlyTyrSerHisSerCysCys 560
 1834 TACGTAAATCTCGAGAGAGGTACAGAGCTGACTGACCTGGCTACTCATTTCTTGGCTGC 1893
 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
 1894 ATCACTCAGCACTGCACTTCTTTATAGTAAGTAACTAAGAACCAAGAAATCCACACGT 1953
 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysValThrLysGluPhe 600
 1954 GTGTCCCTTTCAACAGCTATCAAGTCTTGAAGTGAACCAACTGCAAAACAAAGGAATTT 2013
 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePhe 620

2014 TGGGCCACCATTTTGGATTCAACAGGTCCTTCCTCGACTATATACCTCCCAAAATTTTC 2073
 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
 2074 TCTTTGAAAGTACTACTGATTTACATTTGATGATGATGCTCTCAAGGCTCATGATCTA 2133
 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
 2134 CAGCTCGAAAGAAATATCTACTGTGCTGTCATATATATGATGCTTCAAGGTGACGTTC 2193
 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
 2194 GTCAATATACGTTTAAAGAGTCAAGTATTCCTGATATACCTAGCTCTTAGGT 2253
 681 TyrValValValValIleAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGly 700
 2254 TATGTGTGTATGATGATGATCAACAGGGATCTCTGACCGAGGCTTAAATTTGAAGGC 2213
 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
 2314 GCCTTAAATATTAATATGGTCAAAATGAATTAAGATCAAGTCAAGTCAAGTCAATAT 2373
 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyTyrIleHisGlyTrpSerTyr 740
 2374 CTAGCTTCTCGATATGATTTTCACTTGAATAGATCGTGGGCAATCCAGGCTGGTCTAT 2433
 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
 2434 GAGAGATACCTCTCCCTGATGGCATTAATGCAGAGTCAAGATATCTTCAGGCTTGCTATT 2493
 761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
 2494 GCTGGGGCCCAAGTCACTGTGGATCTTTAAGATACAGATACAGGAACTATATATG 2553
 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
 2554 GGTCACTTACCAAGATGAAGAGGCTATTACTTATGATCTGTGGCATGCAAGAGAA 2613
 801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
 2614 AAGTCCCTCTGAAACCAATCTGTTACTCTCTTAAATGATGTTCCGAGAGAAAGTTC 2673
 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
 2674 CATTTGCAATACCAAGTATATTTACTAGTTTTTTAGTGAAGGCTGAAAGCCATATGAT 2733
 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
 2734 TTACAGATCTATCTCTGAGAGACACAGCATTAAGATTCCTGAATCGGAGAACATTAAT 2793
 861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
 2794 GAATGATCTCTTTGCACTTCACTTCAAGAAACCTTGATCAAGTATGCTGCTTAA 2853
 880 ValIle 882
 2854 GTGATTA 2855

RESULT 5

AAD38956
 ID AAD38956 standard; cDNA; 3120 BP.

AC AAD38956;

DT 23-SEP-2002 (first entry)

DE Human dipeptidyl peptidase 8 (DPP8) cDNA.

KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
 autoimmunity; human immuno deficiency virus; HIV infection; cyostatic;
 graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
 antiviral; enzyme; gene; ss.

```

XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 214..2862
FT /tag=a
FT /product= "Human DPP8 protein"
XX
XX MO200234900-A1.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001W0-AU0388.
XX
XX 27-OCT-2000; 2000AU-0001078.
XX
XX (UNSY ) UNITV SYDNEY.
XX
XX Abbot CA, Gorrell MD;
XX
XX WPI: 2002-454646/48.
XX
XX P-PSDB; AAE24170.
XX
XX New dipeptidyl peptidase (DPP) peptides, useful for screening
XX inhibitors of DPP catalytic activity, which may be employed to treat
XX e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV infection -
XX
XX Example; Fig 1; 91pp; English.
XX
XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX polynucleotides encoding such proteins. The DPP peptides are useful for
XX screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV (human immunodeficiency virus) infection. The present
XX sequence is human DPP8 cDNA.
XX
SQ Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 3120
Score: 4700.00 Matches: 882
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-10-070-464-1 (1-882) x AAD38956 (1-3120)

QY 1 MetaAlaAlaMetGluThrGluGluLeuGlyValGluIlePheGluThrAlaAspCys 20
Db 214 ATGCACAGCAGCAATGGAACAGAACAGCTGGGTGTTGAGATATTTGAACCTGCGACATGT 273
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
Db 274 GAGAGAGATATTGATACACAGATCGGCTTAATTGGAGCCTTTTATGTTGAACGGAT 333
QY 41 SerTyrSerGluLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db 334 TCCGGAGTCACTTAATAAAAGCTCTGCTCCGATACCGAAATATCATGTGCTACATGATG 393
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db 394 GCTAAGGCACCACTGATTTTCATCTTTGTGAAGAGATGATCCAGATGGACCTCATTTCA 453
QY 81 AspArgGlyLeuTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db 454 GACAGATCTATTACCTTGGCATGTCTGTGAGAAACAGAAATATACAGTTTATCTTCT 513
QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu 120
Db 514 GAAATTCCTCAAACTATCAATAGACAGAGTCTTAATGCTCTCTTGAAGCCCTCTTTTG 573

```

```

QY 121 AspLeuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
Db 574 GATCTTTTTCAGGCAACACTGACATGATGATATTTCTGAGAAAGAACTATTAAAG 633
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlyGlySerGly 160
Db 634 GAAAGAAACGATTTGAAACAGTCCGAATTTGCTTTTACGATTAATCCACAGAAAGTGGCA 693
QY 161 ThrPheLeuPheGluAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
Db 694 ACATTTCTGTTCCAGCCGGTGTGTAATTTATACGTAATAAGATGAGGCGACAGAGA 753
QY 181 PheThrGlnGluProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db 754 TTTACGCAACACCTTTTAAGCCCATCTAGTGAAGAACTAGTTGCTCCAAACATACGATG 813
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
Db 814 GATCCAAATTTATGCCCCGCTGATCCAGACTGATTTGCTTTATACATGACAAAGATATT 873
QY 221 TrpIleSerAsnIleValThrArgGluGluArgValGlyLeuThrTyrValHisAsnGluLeu 240
Db 874 TGGATATCTTAACATGTAACCAAGAAAGAAAGAAAGAACTCATTAATGTGCAAAATGAGCTA 933
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db 934 GCCAACATGGAAGAGATGCCAGATCAGCTGAGTCCGTAACCTTTGTTCTCCAGAGAGA 993
QY 261 PheAspArgTyrSerGlyTyrTyrTrpCysProLysAlaGluThrThrProSerGlyGly 280
Db 994 TTGTGATATATTCGGCTATTTGGTGGTGTCCAAAGCTGAACAACTCCCGATGGTGGT 1053
QY 281 LysIleLeuValGlyLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
Db 1054 AAAATTTCTTAATTTCTATGAGAAATGATGATCTGAGTGGAATTTATTCATGTT 1113
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db 1114 ACATCCCTTATGTTGGAACAAAGAGGCGACAGTTCATCCGTTATCCTTAACAGATACA 1173
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db 1174 GCAATCTTAAGTCACTTTTAAGATGTCAGAAATAAAGATTTGATGCTGGAAGAGAGATC 1233
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db 1234 ATAGATGTCATGATAGAACTAATTTCAACCTTTTGAGATTTCAATTTGAAGAGAGTTGAA 1293
QY 361 TyrIleAlaArgAlaGlyTyrTrpThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
Db 1294 TATATTGTCNAGCTGATGATGATCCTGAGGGAATATGCTGTGTCATCTACTAGAT 1353
QY 381 ArgSerGlnThrArgLeuGluIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db 1354 CGTCCCAAGACTCCCTACATAGATGTGTGATCCACTGATTAATTTATCCCACTAGAA 1413
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db 1414 GATGATGTTATGAAAGCAGAGACGACTGATGATGATGCTGATTTCTGTAGCCGACTA 1473
QY 421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
Db 1474 ATTATCTATAGAAAGAAACAGACATCTGGATTAATATTCATGATCATCTTTCACTTTT 1533
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db 1534 CCCCAGAGTCAAGAGAGAAATTTAGTTTATTTTTCCTCTGATCCAAACAGGTTTC 1593
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
Db 1594 CGTCATTTATACAAATTTACATCTATTTTAAAGAAAGCAAAATTAACGATTCACAGTGGT 1653
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500

```